Analysis of selectively ^{15}N labelled PpiB, proinsulin connecting peptide and DnaG-C by NMR

by

Bogdan Bancia

BSc Iasi-Romania

A thesis submitted for the degree of Master of Philosophy

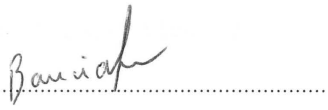
of

The Australian National University

June 2004

Declaration

The work described in this thesis is original and has not previously been submitted for a degree or diploma in any other University or College, and to the best of my knowledge, does not contain material previously published or presented by another person, except where due reference is made in the text.



Bogdan Bancia

Acknowledgement

I would like to thank my supervisors, Professor Gottfried Otting and Dr Max Keniry for their patience and support provided throughout the completion of these projects. I would also like to thank my advisor Dr Nick Dixon for his friendly explanations and for encouraging me to learn and understand more about protein expression. Thankyou also to the other members of my supervisory panel, Dr David Ollis for his interest and offering friendly advice when needed.

Thank you to Dr Kiyoshi Ozawa and Dr Madeleine Headlam for providing the PpiB samples, Karin Loscha for expressing the DnaG-C protein and Dr Neal Williams for teaching me subcloning and overexpression experiments. I would also like to express my appreciation for the efforts and the support provided by Chris Blake with the Varian NMR system, Chris Delfs for computer assistance and Penny Lilley for general assistance in Dixon lab.

Contents

Abstract

The three-dimensional structure determination of proteins represents an important step towards understanding their biological function and thus their roles in living organisms. Using a combination of multidimensional NMR techniques three different biomolecules were analyzed in the present study, *E. coli* peptidyl – prolyl cis-trans isomerase PpiB, proinsulin connecting peptide and DnaG-C.

^{15}N-HSQC spectra were recorded of PpiB which had been expressed without further purification in a cell-free expression system with amino acid selective isotope labelling. Comparison of spectra before and after ultrafiltration indicated that labelled metabolic by-products are of low molecular weight. Therefore, the labelled protein signals are easily distinguished from those of the metabolites.

The structure analysis of the proinsulin connecting peptide included the assignment of ^{1}H, ^{13}C and ^{15}N NMR resonances using 2D NMR, measurement of T_1 (^{1}H) relaxation times and comparison of chemical shifts with sequence dependent chemical shift prediction. The results showed that the peptide assumes a random-coil conformation in aqueous solution.

The NMR resonance assignment of the C-terminal domain of *E. coli* DnaG primase was performed with the goal to assess its structural flexibility in solution. A 2D TOCSY experiment recorded with a long mixing time identified mobile residues at the N- and C-terminal ends. A ^{15}N-HSQC experiment and a number of triple resonance 3D experiments (HNCA, HNCO, CACB(CO)HN, HNCACB, HNCAHA and NOESY-^{15}N-HSQC) were used for the backbone assignment. $\{^{1}H\}^{15}N$ NOEs were measured to characterize the mobility of the protein backbone.

Overall, these projects demonstrate the power of NMR spectroscopy for the study of biomolecules in solution, in particular with respect to the analysis of structural flexibility which presents an important aspect of their biological function.

List of abbreviations

2D	two-dimensional
3D	three-dimensional
CD	circular dichroism
COSY	correlated spectroscopy
DQF	double quantum filtered
FT	Fourier transform
HSQC	heteronuclear single quantum coherence
INEPT	insensitive nuclei enhanced by polarization transfer
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
REBURP	refocusing, band-selective, uniform response, pure-phase
SL	spin-lock
TOCSY	total correlation spectroscopy
WATERGATE	water suppression by gradient tailored excitation
WURST	wideband, uniform rate, smooth truncation

CHAPTER 1

Introduction

Next page

1.1 *Proinsulin connecting peptide*

C-peptide is derived from its precursors, preproinsulin and proinsulin, in the course of insulin biosynthesis. It has been considered to have no physiological effect on its own but recent data demonstrated that it binds with high affinity to cell surfaces, probably to G protein-coupled receptors, with subsequent activation of Ca^{2+}-dependent intracellular signaling pathways and stimulation of Na,K-ATPase activities (**1**). It was further shown that C-peptide is a biologically active peptide hormone with beneficial effects on kidney and nerve function in type 1 diabetes. (**2**).

It is well established that solvents such as trifluoroethanol (TFE, **3**) and dodecylphosphocholine (DPC) micelles (**4**) can induce defined conformations in small peptides which normally exhibit no stable secondary structure in aqueous solution. These mixed solvent systems, because of their dual lipophilic and aqueous nature, can also stabilize structures that more closely approximate the biologically relevant conformations of protein-bound and membrane-bound molecules (**3**). A number of studies suggest that peptides acting on membrane-bound molecules bind to the lipid membrane before interacting with the receptor (**5**). The lipid membrane can assist receptor binding by locking the peptide into its bioactive shape and thereby changing the conformational search from a three dimensional to a two-dimensional process (**6**).

Circular dichroism (CD) experiments have demonstrated that C-peptide exhibits a random coil conformation in water and micelles, whereas the analysis of C-peptide $^{1}H^{\alpha}$ NMR chemical shifts in 95% TFE/ 5% H_2O showed that residues 1-11 adopt a α-helical conformation but no globular three dimensional structure (**7**). Within this project, the ^{1}H, ^{13}C and ^{15}N NMR resonances of the C-peptide have been assigned using a 36 mM sample in 90% H_2O/10% D_2O. NOEs were observed only between neighbouring residues, showing that the C-peptide adopts a random coil conformation in aqueous solution. $T_1(^{1}H)$ relaxation rates are also in agreement with a random coil conformation. The observed chemical shifts were compared to the chemical shifts predicted for random coil peptides (**8**) and to literature values (**9**). The differences between the observed and the predicted $^{1}H^{\alpha}$ chemical shifts were small, supporting the notion of a random coil conformation of the C-peptide in solution.

1.2 DnaG-C

The DnaG primase is a single-stranded DNA-dependent RNA polymerase and is an essential component of the replisome, the machine that accomplishes chromosomal DNA replication in *Escherichia coli* (**10**). It is the enzyme responsible for repeated synthesis of short oligoribonucleotide primers that are extended by DNA polymerase III holoenzyme to produce Okazaki fragments on the lagging strand at the replication fork, and it also synthesizes the initiating RNA primers on both leading strands at the origin of replication (*ori*C). It comprises three domains: a small N-terminal zinc-binding domain, a larger central domain responsible for RNA synthesis, and a C-terminal domain comprising residues 434 – 581 (DnaG-C) that is believed to interact with the hexameric DnaB helicase (**11**). DnaG-C was expressed in *E. coli* under the control of tandem bacteriophage λ promoters and was purified in yields of 4-6 mg/L of culture and studied by X-ray crystallography and NMR (**12**). The linewidths observed in a ^{1}H NMR spectrum of a 0.5 mM sample at pH 6.05, and 36 ^{0}C in the presence of 10 mM phosphate, suggested that the protein is monomeric. DnaG-C is expected to be monomeric in solution since the full-length primase is monomeric under all conditions that have been examined (**13**). Before the X-ray measurements, a 2D TOCSY experiment recorded with 80 ms mixing time identified only a few mobile residues at the N- and C-termini, indicating that the structured core of the protein comprises residues 444-579 (11-146 in our notation). A ^{15}N-HSQC experiment and a number of triple resonance 3D experiments (HNCA, HNCO, CACB(CO)HN, HNCACB, HNCAHA and NOESY-^{15}N-HSQC) were recorded using uniformly ^{13}C/^{15}N-labelled DnaG-C and 136 of 141 possible backbone amide resonances could be assigned by the sequential assignment procedure. {^{1}H}^{15}N NOEs measured to characterize the mobility of the protein backbone showed that the backbone of DnaG-C is rigid, at least for the assigned residues.

1.3 NMR approach to protein structure determination

1.3.1 Introduction

Nuclear magnetic resonance in bulk matter was observed for the first time in 1946 by two groups of researchers separately led by Felix Bloch and Edward Purcell (Nobel Prize in Physics, 1952). Soon, it was shown that the NMR signal carried information

about the chemical environment of the nucleus studied (**14**). This prompted the intensive development of NMR as an analytical tool in chemistry, especially after the discovery of spin-spin coupling arising from interactions between nuclei sharing bonding electrons (**15**). Another important contribution was the discovery of the Overhauser effect (**16**) where the intensity of a NMR signal is enhanced by the irradiation of unpaired electron spins in a sample. This lead to the nuclear Overhauser experiment used to determine mutual dipolar interactions of protons in close proximity, thus giving distance information. The potential sensitivity of NMR was substantially increased with the development of pulse methods with Fourier transformation (Ernst, 1966; Nobel Prize in 1991; **17**) and two-dimensional NMR techniques (**18**).

The role of NMR in structural biology was established when the first two-dimensional spectrum of a protein was recorded in 1977 by Nagayama (**19**). In the early 1980s, Kurt Wüthrich (Nobel Prize in 2002; **20**, **21**, **22**) demonstrated that the ^{1}H resonances in proteins could be assigned by a method called sequential assignment. In 1985 Wüthrich and coworkers published the first complete three-dimensional structure of a small protein in solution, primase inhibitor IIA from bull seminal plasma, BUSI IIA (**23**). Since that time, high resolution NMR has become an important tool for studying the structure, dynamics and molecular interactions of biomolecules, an alternative and complementary method to X-ray crystallography. Of the total 14,734 atomic coordinate sets deposited in Protein Data Bank up to 2002, about 20% had been determined by NMR. Improved sensitivity and resolution were achieved using more powerful magnets with spectrometers operating at ^{1}H frequencies as high as 900 MHz. Sophisticated pulse schemes were developed to record triple resonance 3D spectra on selectively and uniformly isotope labelled proteins, enhancing the utility of NMR for the study of protein structure and dynamics (**24**). The ^{15}N or ^{13}C labels are incorporated into the proteins by modern biotechnology methods to replace the naturally most abundant ^{14}N and ^{12}C nuclei which have unfavourable magnetic properties.

Spectral quality was substantially enhanced with the introduction of pulsed field gradient techniques for artefact and solvent suppression (**25**). Artefacts in NMR were initially eliminated by phase cycling when the experiment is repeated several times

with different phases for the RF pulses and receiver to ensure that the desired coherence transfer pathway is retained while the unwanted coherence transfer pathways cancel at the end of the phase cycle. An experiment employing phase cycle artefact suppression often requires hours to days to be recorded, thus requiring high spectrometer stability for all this period of time. The use of pulsed field gradients to refocus the coherences of the desired pathway and dephase the unwanted coherence pathways permits artefact suppression on a per scan basis, thereby reducing the need for spectrometer stability (**26**, **27**).

The solution structure of a protein is determined by translating distance and dihedral angle information into atomic coordinates. This process follows several well-established steps. The first step is the sequence specific assignments of NMR resonances. The next step is the translation of structure related parameters (J coupling data and NOE data) into dihedral angle restraints and interproton distance restraints. Finally, using computational techniques, a 3D structure is calculated from these dihedral angle and interproton restraints. The strategy required to sequentially assign resonances differs depending on whether an unlabelled or labelled protein is available. For the assignment of C-peptide, two-dimensional homo- and heteronuclear methods were used. Heteronuclear spectroscopy partially resolves the chemical shift degeneracy in two-dimensional homonuclear spectra by dispersing the resonances along a better resolved heteronuclear frequency axis (^{15}N or ^{13}C). In these experiments, magnetization is transferred via conformationally independent one bond-coupling constants using INEPT modules (**28**).

1.3.2 Assignment of homonuclear and heteronuclear 2D spectra

As mentioned above, resonance assignment in multidimensional NMR spectra follows a procedure called sequential assignment. The first step in this procedure is the identification of characteristic spin system patterns by 2D COSY and TOCSY experiments. In these experiments, the connectivities to the amide protons are confined to intraresidue contacts because no ^{1}H J-coupling is observed across the peptide bond. The spin systems can be divided into four families: short sidechain (Gly, Ala and Thr), long sidechain (Val, Ile, Leu), type J (Ser, Asp, Asn, Cys, Trp, Phe, Tyr and His) and type U (Lys, Arg, Met, Gln and Pro) and are assigned based on

coupling patterns and chemical shifts. Some amino acids have easily recognizable spin system patterns and chemical shifts that uniquely identify that amino acid. For example, glycine is the only amino acid that has two alpha protons and is readily identified. Valine, leucine and isoleucine can be recognized by their two methyl groups which give a characteristic row of double signals between 0 and 1.5 ppm. Similarly, alanine and threonine are identified by their single methyl resonances. Type J residues have $^1H^{\beta}$ resonances downfield of 2.5 ppm and have identical $^1H^N$-$^1H^{\alpha}$-$^1H^{\beta}$ spin subsystems. The aromatic protons of Tyr, Phe, Trp and His residues and sidechain amide protons of Gln and Asn are not scalar coupled to the remainder of the sidechain, association of these resonances being possible only on the basis of intraresidue NOE correlations.

In the second step of the assignment process, the sequential NOE contacts from 2D NOESY spectra are used to connect the previously assigned amino acids to the neighbouring amino acids in the primary sequence.

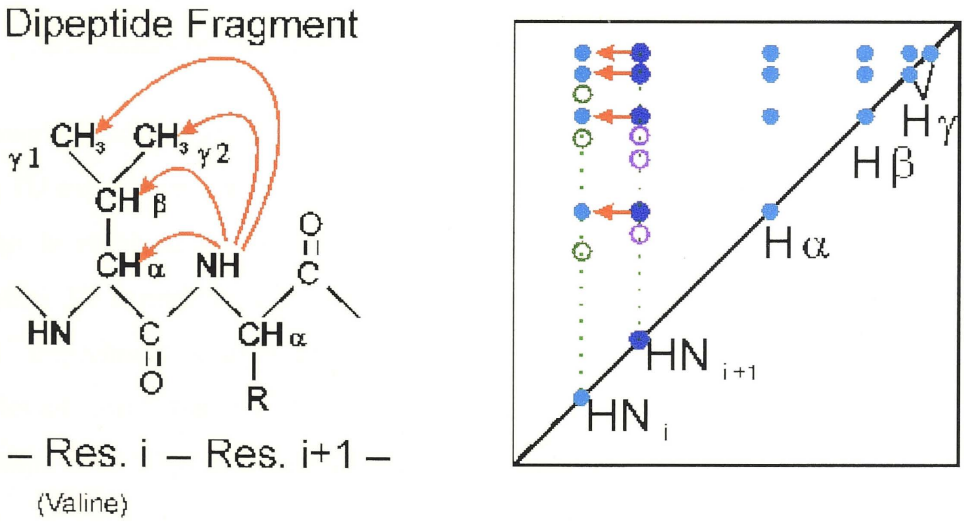


Fig. 1 Dipeptide unit showing the sequential NOE connectivities. d_{NN}, $d_{\alpha N}$, $d_{\beta N}$ and $d_{\gamma N}$ used for sequential proton assignments.

The connectivity of a given amino acid in the sequence (i) to its following one (i+1) can be observed in the NOESY spectra because the distance of the $^{1}H^{N}$ (i+1) proton to the α, β and γ protons of (i) is smaller than 5 Å in almost all cases. Therefore, sequential cross peaks of α(i), β(i) etc (Fig. 1, blue) are observed at the frequency of the $^{1}H^{N}$ (i+1) proton. These inter-residual cross-peak signals were distinguished from the intra-residual ones by comparing the 2D NOESY and TOCSY spectra. Proline residues interrupt the chain of sequential connectivities because they lack $^{1}H^{N}$ protons. Heteronuclear Single Quantum Coherence (HSQC) spectra and the analogous 2D experiments were used to assign ^{13}C and ^{15}N resonances and to resolve the overlap in the homonuclear spectra by expanding the resonances in a third ^{13}C or ^{15}N dimension. For example, a ^{1}H-^{15}N (or ^{1}H-^{13}C) HSQC experiment can be combined with a TOCSY experiment to yield a TOCSY-relayed ^{15}N (or ^{13}C)-HSQC experiment.

In this experiment, a TOCSY pulse sequence is inserted before the acquisition period of the HSQC experiment and TOCSY cross peaks for all protons are connected to the $^{1}H^{N}$ proton for which the signal was acquired, allowing separation of overlapping resonances.

1.3.3 Assignment of 3D triple resonance spectra

The overlap and the ambiguities in the interpretation of 2D spectra can be solved by using 3D experiments in combination with ^{15}N- or/and ^{13}C-labelling procedures. The overlap is resolved by spreading the resonances in a new dimension. The most important 3D experiments used for backbone resonance assignment consist of a ^{15}N-HSQC experiment with a third ^{13}C or ^{1}H dimension. For example, the HNCA experiment correlates the $^{1}H^{N}$ and ^{15}N resonances, with the intraresidual (i) and sequential (i-1) $^{13}C_{\alpha}$ chemical shifts (Fig. 2).

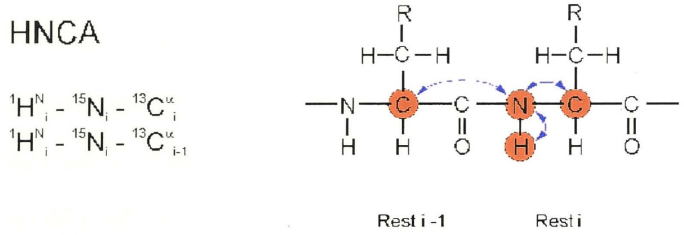


Fig. 2 The 3D HNCA experiment correlates $^{1}H^{N}$ and ^{15}N chemical shifts with the $^{13}C^{\alpha}$ of the same residue (i) via one-bond coupling constant $^{1}J_{NC}$ (~15 Hz) and with the $^{13}C^{\alpha}$ of the preceding residue (i-1) via the two-bond connectivity $^{2}J_{NC}$ (~7 Hz).

The backbone resonance assignment starts with the acquisition of a ^{15}N HSQC spectrum, which provides correlations between the $^{1}H^{N}(i)$ and $^{15}N(i)$ nuclei of each residue (i). Starting from the $^{1}H^{N}(i)$ and $^{15}N(i)$ correlations, the $^{13}C^{\beta}$, $^{13}C^{\alpha}$, ^{13}CO and ^{1}H resonances of the preceding residue are established using HNCA, HNCO, CBCA(CO)NH, HNCAHA and NOESY-^{15}N-HSQC experiments. Table 1 shows that the backbone resonance assignment may be accomplished by matching of four intraresidue backbone resonances of residue (i) to four sequential resonances of the preceding residue (i-1).

	(i-1)				**(i)**		**(i)**			
	C^{β}	**C^{α}**	**CO**	**H^{α}**	**N**	**NH**	**C^{α}**	**C^{β}**	**CO**	**H^{α}**
HSQC					**x**	**x**				
HNCA		(+)			x	x	+			
HNCO			+		x	x				
CACB(CO)HN	+	+			x	x				
HNCACB	(+)	(+)			x	x	+	+		
HNCAHA				(+)	x	x				+
NOESY-HSQC	Secondary structure elements				x	x	d_{NN}, $d_{\alpha N}$, etc			

Table 1 Connectivities observed in 3D triple-resonance experiments used for the sequential assignment of DnaG-C; (+) sequential connectivities

Sorting the $^{1}H^{N}(i)$-$^{15}N(i)$ connectivities by finding amino acids with a characteristic pattern of cross peaks (i.e. Gly, Ala, Thr and Ser from the ^{13}C dimension of the CBCA(CO)HN or HNCACB experiments) results in a number of peptide fragments, which have to be matched with the protein sequence. For example, glycine is the only amino acid that has no β-carbon, a characteristic upfield C^{α} chemical shift at about 46 ppm which is readily identified in the CBCA(CO)HN spectrum, alanine has a β-carbon resonating at about 20 ppm, threonine and serine residues can be easily recognized by their characteristic β-carbon chemical shifts (60-70) ppm. Once the backbone resonances are assigned by this "through-bond" sequential assignment procedure, the secondary structure elements are easily identified from characteristic medium-range intraresidue NOEs in NOESY-^{15}N-HSQC spectra.

1.3.4 Homonuclear experiments

1.3.4.1 Introduction

A general scheme for recording homonuclear two-dimensional spectra can be divided into the four distinct parts illustrated in Fig. 3.

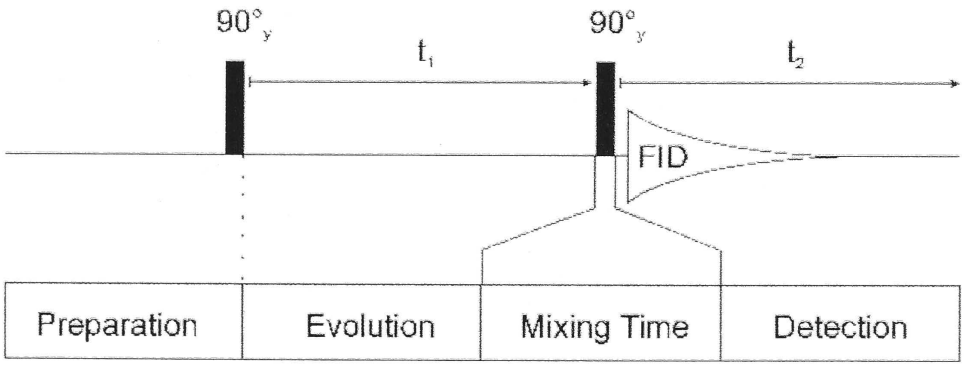


Fig. 3 General scheme for two-dimensional homonuclear spectroscopy. It can be divided into four defined components: preparation, evolution, mixing period and detection.

During the preparation stage a single quantum coherence is created using a 90^{0} pulse applied after a relaxation delay. The evolution period contains a variable time delay that is incremented during the course of a two-dimensional NMR experiment (frequency labelling in F_1) and for each incrementable delay the pulse sequence is executed a few times, the signals being co-added for signal averaging or phase

cycling. The key to obtaining useful chemical or structural information from two-dimensional NMR spectroscopy is the transfer of coherence from one spin to another during the mixing period: via through-bond scalar coupling (J correlated spectroscopy, COSY and TOCSY) and via through-space dipolar interactions (nuclear Overhauser effect spectroscopy, NOESY). Finally, the transverse magnetization is monitored during the acquisition period.

1.3.4.2 DQF-COSY and TOCSY ('through-bond connectivities')

In general, J-correlated spectroscopy (COSY) allows us the determination of the connectivity of protons in a molecule by transferring magnetization from one spin to another via J-coupling (through-bond transfer). The DQF (double quantum filtered) – COSY (**29**) experiment consists of the 90° - t_1 – 90° element of COSY plus an extra 90° pulse applied immediately after it. In this experiment, the dispersive diagonal peaks of the conventional COSY experiment are largely suppressed; Fig. 4). The final 90_x° pulse induces the transitions $-I_xS_y \rightarrow I_xS_z$ (antiphase J-modulated precession with frequency Ω_I during the detection time t_2 giving rise to diagonal peaks) and $-I_yS_x \rightarrow I_zS_y$ (antiphase J-modulated precession with frequency Ω_S, which yields the cross peaks).

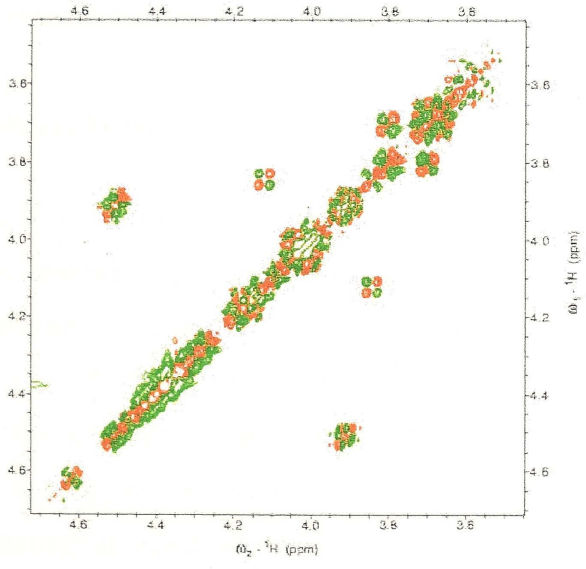


Fig. 4 Region of the DQF-COSY spectrum recorded on C-peptide in 90% H_2O/10% D_2O, at pH 6 and 25 °C, showing the antiphase absorption mode of the cross peaks. The experiment provides correlations between protons separated by two and three bonds.

A protein consists of a string of amino acid residues connected by amide bonds and since there is no detectable ^1H-^1H coupling across the amide bond, each amino acid residue can be considered an independent spin system which is readily identified using J-correlated spectroscopy (COSY or one of its modified versions). The region containing the $^1H^N$ to $^1H^\alpha$ cross peaks of the DQF-COSY spectrum is known as the "fingerprint region" as it is generally well dispersed and characteristically different for each protein. Each cross-peak which has a characteristic antiphase pattern from the $^1H^N$-$^1H^\alpha$ J-coupling, belongs to a different residue in the sequence.

The TOCSY (total correlation spectroscopy) experiment is similar to COSY, in that it maps out which protons are J-coupled to each other. In the TOCSY spectrum, however, correlations are seen between all protons in a spin system (**30**). The experiment starts with a 90° pulse creating transverse magnetization followed by an incrementable period (t_1) to yield the indirect F_1 dimension after 2D Fourier transform (FT). Between the two times t_1 and t_2 the coherence transfer is performed using a spin-lock period consisting of a train of 180° pulses. During the spin-lock (SL) the in-phase coherence of a spin is transferred into the in-phase coherence of its J-coupled partner, and back, in a oscillatory manner. The additional 90° pulses at the end of t_1 and beginning of t_2 are needed to convert magnetization components into I_z and then back into detectable magnetization.

1.3.4.3 NOESY ('through-space connectivities')

The purpose of the NOESY experiment is to establish connectivities between spins via cross-relaxation. Instead of the transfer of antiphase coherence, I_yS_z -> I_zS_y (the key process of correlated spectroscopy), NOESY relies on transfer of longitudinal magnetization (I_z -> S_z). The experiment starts with a 90° pulse to bring magnetization into the transverse plane, and then the spins are labelled with their offset frequencies during an evolution period, t_1. A second 90° pulse creates longitudinal z-magnetization again which is transferred from one spin to another via cross relaxation during the mixing time τ_{mix}. The phase cycle is chosen so that at the start of the mixing period only the longitudinal terms are present. A third 90° pulse brings magnetization back into the transverse plane, the Larmor frequency of which is monitored during the acquisition period t_2. The relative sign of the cross peaks

depends on the rotational correlation times of the molecules. For long rotational correlation times, the two sets of spins are in a state of internal equilibrium in which the magnetization of the two species become equal and decay together leading to cross-peaks and diagonal peaks of the same sign. For short correlation times, the sense of magnetization transfer during the mixing time is negative (magnetization of the two spins have opposite signs) leading to cross-peaks and diagonal peaks of opposite signs.

For short mixing periods the amplitude of the cross peaks is given by:

$$a_{cross} = R_{cross}\tau_{mix} \quad (1)$$

where R_{cross} is the cross relaxation rate which is proportional to the square of the dipole-dipole interaction and hence to the inverse sixth power of the distance between the spins:

$$a_{cross} \sim r^{-6} \quad (2)$$

This distance dependence is exploited for calculating protein structures by measuring the cross peak amplitudes in a NOESY spectrum where a very large number of internuclear distances can be measured. In order to determine the proportionality constant, the distance scale can be calibrated by using proton pairs separated by known distances in rigid parts of the molecule (i.e. the distance between ortho and para protons in an aromatic ring). Unknown distances, r_u, are then calculated from the corresponding peak intensities according to:

$$r_u = r_{cal}\,(NOE_{cal}/NOE_u)^{1/6} \quad (3)$$

Fig. 5 shows a NOESY spectrum recorded on DnaG-C with 100 ms mixing time.

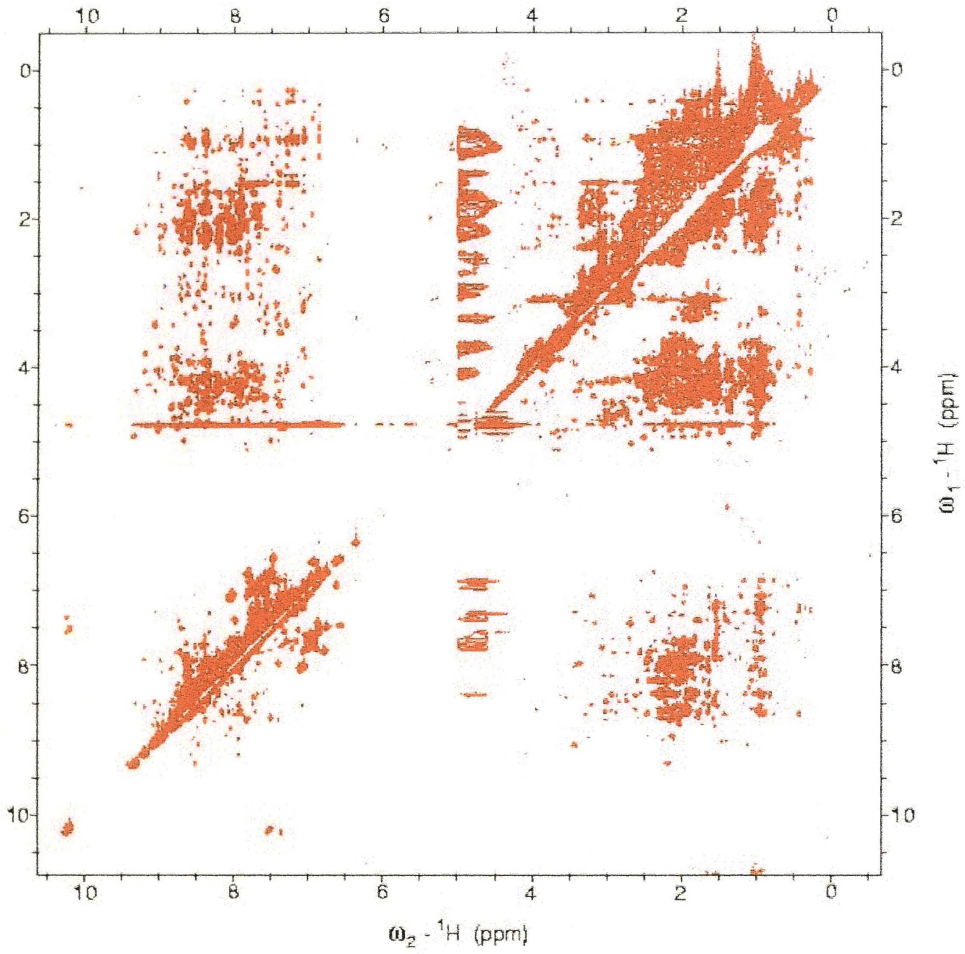


Fig. 5 2D ^{1}H-^{1}H NOESY spectrum recorded on DnaG-C in 90% H_2O/10% D_2O, 100 mM NaCl, 10 mM sodium phosphate, at 36 °C. Each cross-peak can be used to measure an interproton distance equivalent to a structural constraint. 3D structures are determined by using computational methods which search for molecular structures that satisfy the measured NMR constraints.

Often, signal overlap limits the effective resolution making the sequential assignment a very difficult task. Broadband decoupling techniques can be used to decouple certain spectral regions (i.e. $^1H^{\alpha}$ resonances) during signal detection (**31**). A number of different schemes were described which achieve decoupling in the indirectly detected F_1 dimension, including constant time experiments (**32**), time reversal sequences (**33**), and the application of a shaped semiselective refocusing pulse in combination with a non-selective refocusing pulse in the middle of the evolution period. In our studies, a WURST sequence was used to irradiate the $^1H^{\alpha}$ resonances of

the C-peptide without affecting the rest of the spectrum, while F_1 decoupling was achieved using a REBURP pulse (**34**) applied in the middle of the evolution period.

Broadband WURST decoupling employs a linear frequency sweep and a constant radio frequency level (**31**). Fig. 6 (A) shows a spectral region of the 2D NOESY spectrum recorded on C-peptide in 90% H_2O/10% D_2O. The resonances of backbone protons are in a spectral region between 7.8 – 8.7 ppm, each of them being a doublet due to scalar coupling with the α-protons, the signals of which are between 3.8 – 4.7 ppm. Fig. 6 (B) and (C) show the same spectral region with decoupling applied to the spectral region of the α-protons during the acquisition time (B) and during the evolution and detection periods (C). The signals are narrower due to the decoupling.

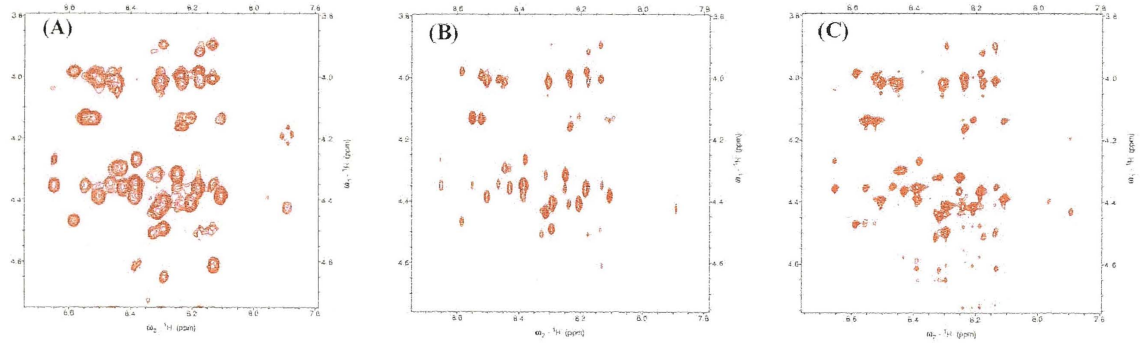


Fig. 6 Spectral region (the region ω_2 = 7.8 – 8.7 ppm and ω_1 = 3.8 – 4.7 ppm) of the 2D NOESY experiment with (B, C) and without (A) α-proton decoupling.

1.3.5 Heteronuclear NMR experiments

1.3.5.1 Introduction

With increasing molecular weights, the linewidths of NMR resonances increase due to increased T_2 relaxation caused by slower molecular tumbling. Therefore the J-correlations involved in the analysis of spin systems become more difficult to observe. At the same time, the number of interactions and therefore the number of cross-peaks in the 2D spectra increase leading to overlap and ambiguities in the analysis of the spectra.

Uniform incorporation of ^{15}N and ^{13}C isotopes into the biological macromolecules combined with 3D heteronuclear pulse schemes to transfer magnetization between scalar (through-bond) and dipolar (through-space) coupled spins, provides access to structural studies of larger proteins in solution. The overlap can be resolved by spreading the resonances in a new spectral dimension and the backbone resonance assignments can be obtained by combining data from a standard set of 3D triple-resonance experiments, all of them involving $^{1}H^{N}$, ^{15}N, $^{13}C^{\beta}$, $^{13}C^{\alpha}$ and ^{13}CO correlations characterized by a large spectral dispersion. The initial procedure proposed by Ikura et al. (**35**) made use of the following 3D experiments: HNCO, HNCA, HCACO and ^{15}N-edited TOCSY-HSQC. The ^{15}N-edited TOCSY-HSQC experiment correlates side chain $^{1}H(i)$ chemical shifts with backbone ($^{15}N(i)$, $^{1}HN(i)$) pairs and in combination with HNCO, HNCA and HCACO enables a complete assignment of $^{1}H^{\alpha}(i-1)$, $^{13}C^{\alpha}(i-1)$, $^{13}CO(i-1)$ and $^{13}C^{\alpha}(i)$ chemical shifts, providing sufficient information to trace the sequential assignment along the polypeptide chain of proteins. For the backbone resonance assignments of DnaG-C, we used the following set of 3D triple resonance experiments: HNCA, CBCA(CO)NH, HNCACB, HNCO, HNCAHA and ^{15}N-edited NOESY-HSQC. Resonances obtained from different 3D spectra are not influenced by isotope shifts leading to a straightforward sorting algorithm, and sometimes even to an automatic procedure (**36**).

1.3.5.2 ^{15}N-HSQC experiment

The ^{15}N-HSQC experiment (**37**) correlates each ^{15}N nucleus and its directly bonded proton spin. Since there is only one backbone amide proton per residue, this experiment provides an important test for the feasibility of an NMR structure determination approach. Low quality spectra indicate that improved sample conditions are required before any further data collection. The coherence pathway can be described as:

$$\mathbf{H^{N}_{i}} \_\_J_{NH}\_\_ \mathbf{^{15}N_{i}\,(t_1)} \_\_J_{NH}\_\_ \mathbf{H^{N}_{i}(t_2)} \qquad (4)$$

where J_{NH} is the one bond N-NH coupling constant, and t_1 and t_2 are the acquisition times.

Since a backbone amide group is represented by a cross peak in the ^{15}N-HSQC spectrum, an HSQC spectrum is often used to monitor ligand binding. Changes or the

disappearance of the signals upon addition of the ligand provide site-specific information about ligand binding (**38**). With prior knowledge about the potential binding site, selective labelling techniques can be used to avoid the assignment of the entire ^{15}N-HSQC spectrum (**39**). The ^{15}N-HSQC spectrum in Fig. 7 shows the cross-peaks for all five methionine residues in ^{15}N-Met labelled PpiB (the *Escherichia coli* peptidyl-prolyl-cis-trans isomerase PpiB). However, the *E. coli* extracts contain enzymes that transform a fraction of the amino acids into low molecular weight metabolites. These compounds give rise to prominent signals in the ^{15}N-HSQC spectrum but are easily removed by ultrafiltration.

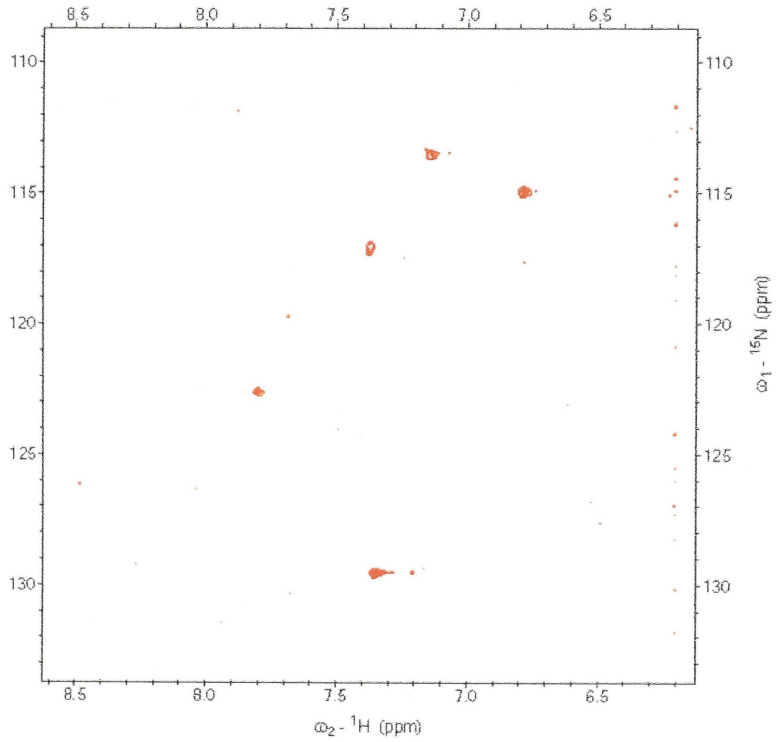


Fig. 7 ^{15}N-HSQC spectrum of PpiB labelled with ^{15}N-Met recorded after ultrafiltration and addition of D_2O 10% (v/v).

1.3.5.3 ^{15}N-edited TOCSY-HSQC and ^{15}N-edited NOESY-HSQC experiments

For proteins with well resolved ^{15}N-HSQC spectra, ^{15}N (or ^{13}C) edited experiments can be sufficient for the complete assignment of the backbone $^{1}H^{N}$ and ^{15}N chemical resonances (**40**). The ^{15}N-edited TOCSY-HSQC experiment provides intra-residue correlations by linking ^{1}H side-chain chemical shifts with backbone (^{15}N, $^{1}H^{N}$) pairs. The ^{15}N-edited NOESY-HSQC experiment, on the other hand, provides intra- and inter-residue correlations linking (^{15}N, $^{1}H^{N}$) resonances to protons within a proximity of ~5 Å (**40**).

The development of 3D ^{15}N-edited spectroscopy in the late 1980s had a major impact in the assignment process and structure determination. The method that is employed is identical to that used when analysing 2D homonuclear spectra (**9**) except that the separation of overlapping strips of cross peaks via the ^{15}N shift facilitates the assignment process.

The magnetization transfer pathways in the ^{15}N-edited TOCSY-HSQC experiment can be described as:

$$\mathbf{^{1}H_{i}(t_{1})}\_\_\mathrm{TOCSY}\_\_\mathbf{^{1}H^{N}_{i}}\_\_J_{NH}\_\_\mathbf{^{15}N_{i}(t_{2})}\_\_J_{NH}\_\_\mathbf{^{1}H^{N}_{i}(t_{3})} \quad (5)$$

where $^{1}H_{i}$ includes all side chain protons as well as the backbone proton NH_{i}, while the flow of magnetization in the ^{15}N-edited NOESY-HSQC experiment is:

$$\mathbf{^{1}H_{j}(t_{1})}\_\_\mathrm{NOESY}\_\_\mathbf{^{1}H^{N}_{i}}\_\_J_{NH}\_\_\mathbf{^{15}N_{i}(t_{2})}\_\_J_{NH}\_\_\mathbf{^{1}H^{N}_{i}(t_{3})} \quad (6)$$

with $^{1}H_{j}$ and NH_{i} within a proximity of ~5 Å, so that magnetization transfer via the nuclear Overhauser effect (NOE) can occur between these species. Strip plots from 100 ms ^{15}N-edited NOESY-HSQC spectrum of ^{15}N/^{13}C-labelled DnaG-C (0.6 mM, 90% H_2O/10% D_2O, pH 6, 100 mM NaCl, 10 mM sodium phosphate, 36 ^{0}C) are shown in Fig. 27B.

The experiments can also be recorded in a 2D manner; the coherence pathways being described as:

$$\mathbf{^{1}H^{N}_{i}}\_\_J_{NH}\_\_\mathbf{^{15}N_{i}(t_{1})}\_\_\mathrm{TOCSY}\_\_\mathbf{^{1}H_{i}(t_{2})} \quad (7)$$

$$\mathbf{^{1}H^{N}_{i}}\_\_J_{NH}\_\_\mathbf{^{15}N_{i}(t_{1})}\_\_\mathrm{NOESY}\_\_\mathbf{^{1}H_{j}(t_{2})} \quad (8)$$

In the 2D NOE or TOCSY-relayed ^{15}N-HSQC experiments, NOESY and TOCSY cross-peaks are observed at the frequency of each amide nitrogen making them crucial

experiments for the resonance assignment of random coil peptides where many resonances overlap in the proton dimension (Fig. 8). In this situation, resolving the NOESY spectrum along the heteronuclear (^{13}C or ^{15}N) dimension yields an important gain in resolution compared to the conventional NOESY experiment, while the number of peaks remains constant.

For protein structure determination, ^{1}H-^{1}H distances are the most important parameter and are derived from NOE intensities. Cross peaks in these experiments do not only depend on the cross-relaxation rate, but also on the transfer amplitude of the heteronuclear correlation part. For an accurate estimation of the ^{1}H-^{1}H distances, a HSQC spectrum recorded under the same conditions has to be used in order to correct for the individual transfer efficiencies in the heteronuclear correlation part. This is not needed if NOE-derived distance constraints are only classified according to weak, medium and strong intensities. For proteins with more than 100 amino acids the overlap becomes a problem even in the NOE-relayed ^{15}N-HSQC experiment and 3D triple resonance (^{15}N, ^{13}C and ^{1}H correlations) experiments have to be used.

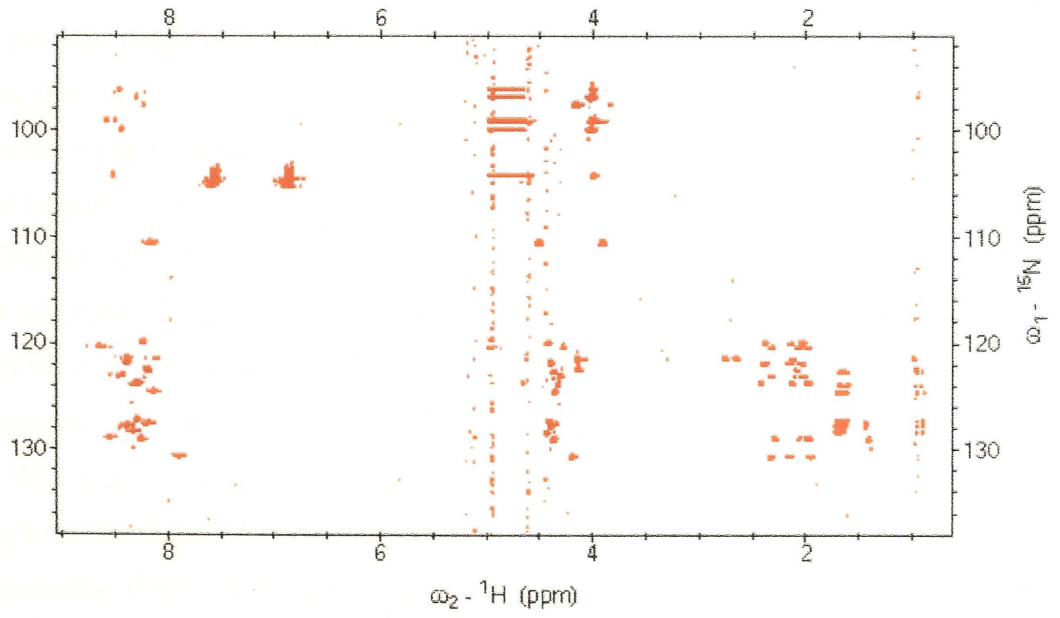


Fig. 8 2D TOCSY-relayed ^{15}N-HSQC spectrum recorded on a 36 mM sample of C-peptide at natural abundance at pH 6 and 25 °C.

1.3.5.4 Triple resonance NMR experiments for backbone assignment

Triple resonance NMR makes use of the J-couplings that exist between the ^{13}C and ^{15}N nuclei and their directly attached protons for the efficient transfer of magnetization, resulting in considerably increased resolution (Fig 9).

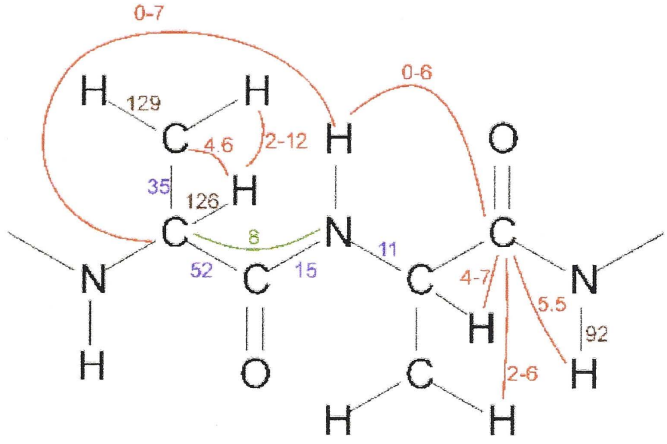


Fig. 9 Spin system of the peptide backbone and the magnitude of the ^{1}J, ^{2}J and ^{3}J coupling constants that are used for magnetization transfer in $^{13}C/^{15}N$ labelled proteins.

For proteins larger than approximately 10 kDa, triple resonance NMR spectroscopy is often the method of choice for backbone assignment and structure determination allowing a significantly larger set of distance constraints to be obtained. A few of the most important triple resonance experiment, which were also used for the assignment of DnaG-C are: HNCACB (**41**), CBCA(CO)NH (**42**), HNCA (**43**), HNCO (**43**) and HNCAHA(**44**). Typically two neighbouring residues are identified through analysis of pairwise complementary 3D experiments which allow one to establish a connection between the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of residue i and residue i-1, and the shifts of the $^{1}H^{N}$ and ^{15}N spins of the ith residue. For example, the HNCACB experiment correlates the (^{15}N, $^{1}H^{N}$) resonances of a particular residue with both the inter- and intra-residue ($^{13}C^{\alpha}$, $^{13}C^{\beta}$) pairs, while the CBCA(CO)NH experiment correlates (^{15}N, $^{1}H^{N}$) resonances with only the inter-residue ($^{13}C^{\alpha}$, $^{13}C^{\beta}$) pairs (Fig. 10).

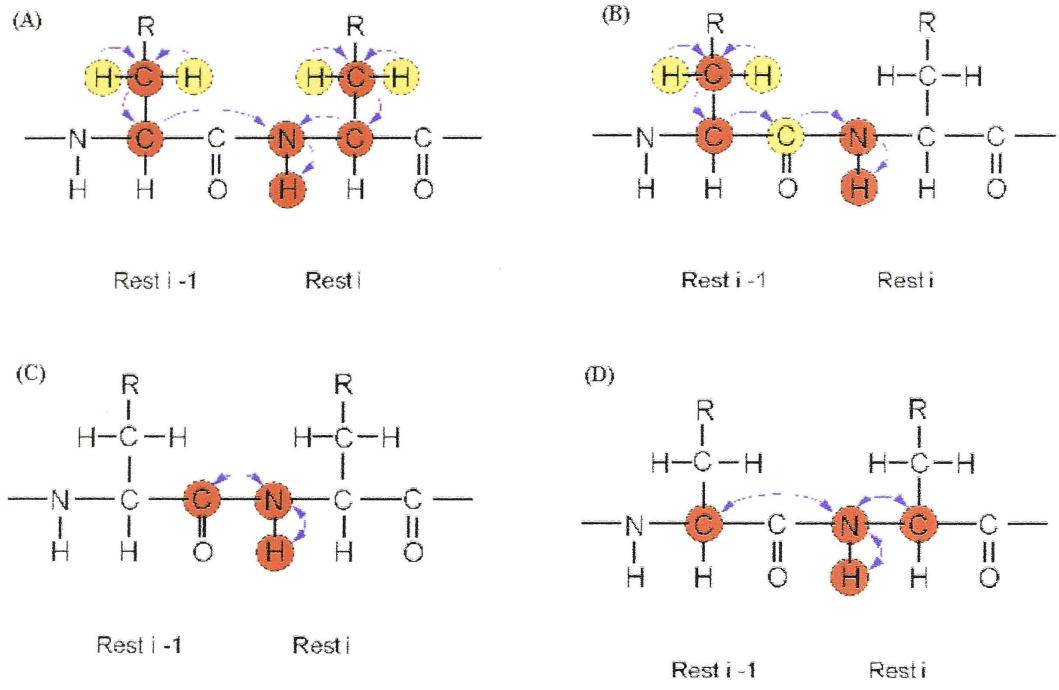


Fig. 10 Chemical shift assignment for $^{13}C/^{15}N$ labelled proteins based on J-correlations (blue lines indicate the connectivities established in the respective experiments). (A) HNCACB, (B) CBCA(CO)NH, (C) HNCO and (D) HNCA experiments.

In the absence of overlap and missing correlations, the HNCACB experiment enables a complete assignment of the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{15}N and $^{1}H^{N}$ resonances. The CBCA(CO)NH experiment is used to identify missing sequential cross-peaks (Fig. 11)

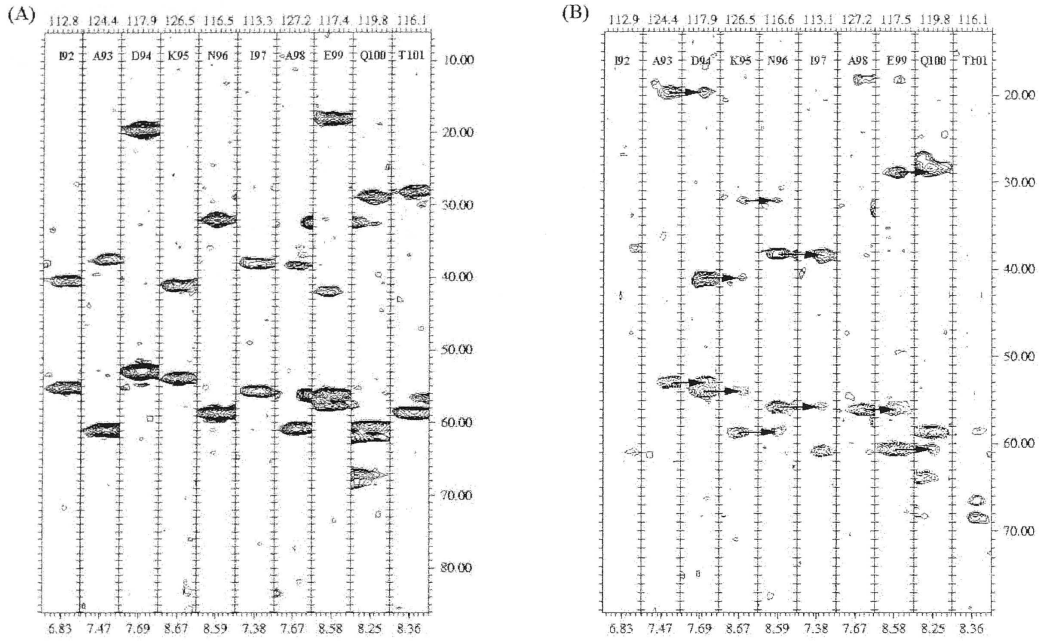


Fig. 11 F_1 ($^{13}C^{\beta}$ and $^{13}C^{\alpha}$) strip plots from CBCA(CO)NH (A) and HNCACB (B) spectra recorded on a 0.6 mM $^{13}C/^{15}N$ labelled DnaG-C sample (90% H_2O/10% D_2O, pH 6, 100 mM NaCl, 10 mM sodium phosphate, 36 °C). Correlations involving amides of residues I92 to T101 are illustrated.

As the size of the proteins increases, it is often necessary to record additional triple resonance experiments to establish unambiguous connectivities. The HNCO experiment (Fig. 12) correlates the (^{15}N, $^{1}H^{N}$) resonances of a particular residue with ^{13}CO of the preceding residue. Because of the relatively large coupling between the amide nitrogen and the backbone carbonyl ($J_{NC'}$ = 15 Hz), this experiment has high sensitivity and is often useful to record prior to any of the other 3D data experiments.

The HNCA experiment is identical with the HNCO, except for the interchange of $^{13}C^{\alpha}$ and ^{13}CO pulses. Two types of cross peaks are observed in the HNCA experiment, namely the correlations $^{1}H^{N}(i)$, $^{15}N(i)$, $^{13}C^{\alpha}(i)$ and $^{1}H^{N}(i)$, $^{15}N(i)$, $^{13}C^{\alpha}(i\text{-}1)$ because both $^{1}J_{C\alpha(i)\ N(i)}$ = 15 Hz and $^{2}J_{C\alpha(i-1)\ N(i)}$ = 7 Hz are of similar size (Fig. 30A). In contrast, only the correlations $N^{H}(i)$, $^{15}N(i)$, $^{13}C'(i\text{-}1)$ are observed in the HNCO spectrum since $^{1}J_{CO(i-1)\ N(i)}$ = 15 Hz, while $^{2}J_{CO(i)\ N(i)} \sim 0$ Hz (Fig. 12). The HNCAHA experiment correlates the (^{15}N, $^{1}H^{N}$) resonances with the intra- and inter-residual $^{1}H^{\alpha}$

resonances providing additional information for the sequential assignment process (**94**).

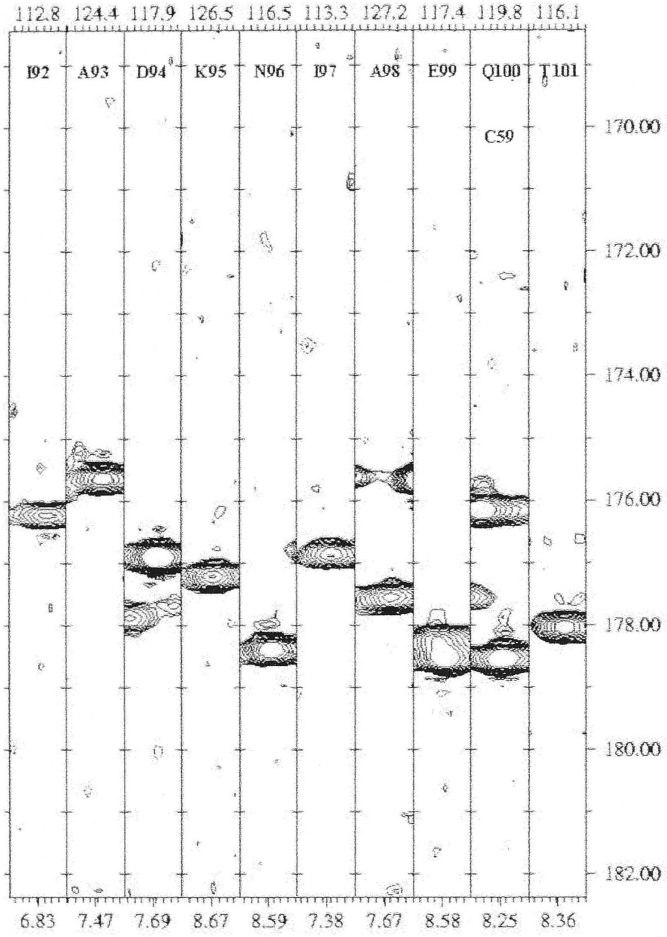


Fig. 12 F_1 strip plots (^{13}CO) from the HNCO spectrum of DnaG-C recorded under the conditions specified above.

All pulse sequences were recorded with sensitivity enhancement in the N-H back transfer and use of constant time evolution of ^{15}N nuclei. Pulsed field gradients were used for solvent and artefact suppression. The coherence pathways can be described as (for simplicity couplings are not indicated):

$$\text{(HNCACB)}\ {}^{1}H^{N} \rightarrow {}^{15}N \rightarrow {}^{13}C^{\alpha} \rightarrow {}^{13}C^{\alpha,\beta}(t_1) \rightarrow {}^{13}C^{\alpha} \rightarrow {}^{15}N(t_2) \rightarrow {}^{1}H^{N}(t_3) \quad (9)$$

$$\text{(CBCA(CO)NH)}\ {}^{1}H^{\alpha,\beta} \rightarrow {}^{13}C^{\alpha,\beta}(t_1) \rightarrow {}^{13}C^{\alpha} \rightarrow {}^{13}CO \rightarrow {}^{15}N(t_2) \rightarrow {}^{1}H^{N}(t_3) \quad (10)$$

$$\text{(HNCA)}\ {}^{1}H^{N} \rightarrow {}^{15}N \rightarrow {}^{13}C^{\alpha}(t_1) \rightarrow {}^{15}N(t_2) \rightarrow {}^{1}H^{N}(t_3) \quad (11)$$

$$\text{(HNCO)}\ {}^{1}H^{N} \rightarrow {}^{15}N \rightarrow {}^{13}CO(t_1) \rightarrow {}^{15}N(t_2) \rightarrow {}^{1}H^{N}(t_3) \quad (12)$$

$$\text{(HNCAHA)}\ ^{1}H^{N}\_\_^{15}N\_\_^{13}C^{\alpha}\_\_^{1}H^{\alpha}(t_3)\_\_\ ^{13}C^{\alpha}\_\_^{15}N\_\_^{1}H^{N} \qquad (13)$$

As the overall transfer efficiency of any experiment is a function of both the J-couplings involved and relaxation times, the values of the delays have to be optimized properly for each experiment. In general, in heteronuclear multi-dimensional NMR, it is necessary to minimize the evolution delays of spins that relax rapidly or have many homonuclear couplings. In a protein backbone these conditions are very important for the $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ spins, which relax rapidly as a result of their dipolar interactions and have several homonuclear coupling partners. The $^{1}H^{\alpha}$ peaks are broadened due to J-coupling to $^{1}H^{N}$ and $^{1}H^{\beta}$ spins and the $^{13}C^{\alpha}$ as a result of J-couplings to the aliphatic side chain carbons. (All pulse sequences can be visualized in Appendix I).

1.3.6 Protein dynamics from NMR

1.3.6.1 Introduction

Proteins are dynamic molecules that often undergo conformational changes while performing their specific functions (i.e enzyme reaction, ligand binding). Various NMR experiments can be used to monitor protein movements over a wide range of timescales. Nuclear spin relaxation rate measurements report the internal motion on the fast (ns-ps) and slow (μs-ms) timescales as well as the overall rotational diffusion of the molecule (5-50 ns), whereas rates of magnetization transfer between protons with different chemical shifts and proton exchange report movements of protein domains on the slow timescales (ms-days) (**45**). Therefore, a complete description of the structure of a protein requires an understanding of how the structure changes with time. The ultimate goal of dynamics is to demonstrate how these motions relate to function.

The development of isotope labelling methods in the late 1980s coupled with multi-dimensional experiments facilitated the measurement of heteronuclear relaxation parameters, especially of ^{15}N spins, which monitor the reorientation of ^{15}N-^{1}H bond vectors. These experiments are used to study backbone dynamics as well as side-chain motions of residues containing nitrogen (**46**, **47**, **48**).

At the same time, methods and theory have been developed that allow a quantitative interpretation of the relaxation of $^{13}C^{\alpha}$ and ^{13}CO spins but a complete description of the relaxation properties of these spins is more complicated than for ^{15}N.

In general, relaxation experiments monitor how the orientation of a particular bond vector changes with time (auto-correlation) or describe correlations between different bond vectors (cross-correlation). Both methods are used to obtain an order parameter describing the amplitude of the internal motions (S^2) and an effective internal rate constant (τ_e) for internal dynamics characterizing the subnanosecond internal motions of bond vectors (^{15}N-^{1}H, ^{13}C-^{1}H, ^{13}C-^{13}C) at numerous sites in the protein backbone.

Protein dynamics studies involve measurement of the longitudinal relaxation rate constant, $R_1 = 1/T_1$, the transverse relaxation rate constant, $R_2 = 1/T_2$, and the steady state $\{^{1}H\}^{15}N$ nuclear Overhauser effect, all being sensitive to internal motions on the submicrosecond timescale. These data are analyzed by the model free approach proposed by Lipari and Szabo (**49**, **50**), assuming either isotropic or anisotropic motion.

1.3.6.2 The model free formalism

Lipari and Szabo (**49**, **50**) showed that the information on fast internal motion derived from NMR relaxation data can be completely described by a generalized order parameter, S^2, which is a measure of the degree of spatial restriction of the motion, and an effective correlation time, τ_e, which is a measure of the rate (time scale) of the motion.

The T_1 and T_2 relaxation times and the NOE enhancement factor of an ^{15}N spin are determined by the fluctuations of ^{15}N-^{1}H vectors with respect to the external magnetic field. The observed quantities are determined by Fourier transform (the spectral density) of a time correlation function evaluated at certain frequencies. These fluctuations are caused by the dipole interaction of the ^{15}N nucleus with its attached proton and by chemical shift anisotropy (**51**).

The following expressions were derived from this model (**49**):

$$1/T_1 = d^2[J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H+\omega_N)] + c2J(\omega_N) \quad (14)$$

$$1/T_2 = (1/2)d^2[4J(0) + J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H+\omega_N)] + (1/6)c^2[3J(\omega_N)+ 4J(0)] \quad (15)$$

$$NOE = 1 + (\gamma_N/\gamma_H)d^2[6J(\omega_H+\omega_N)- J(\omega_H-\omega_N)]T_1 \quad (16)$$

The constants d^2 and c^2 are defined as:

$$d^2 = (1/10)\gamma_H^2\gamma_N^2h2/(4\pi^2)<1/r_{NH}^3>^2 \quad (17)$$

$$c^2 = (2/15)\ \gamma_N^2H_o^2(\sigma_{II}-\sigma_\perp) \quad (18)$$

where γ_H and γ_N are the gyromagnetic ratios of the ^1H and ^{15}N nuclei, ω_H and ω_N are the ^1H and ^{15}N Larmor frequencies, r_{NH} is the internuclear ^1H-^{15}N distance, H_o is the magnetic field strength, σ_{II} and $\sigma_\perp$ are the parallel and perpendicular components of the assumed axially symmetrical ^{15}N chemical shielding tensor.

Within this model, a minimum number of ^{15}N relaxation parameters is required to describe the overall tumbling motion of the protein and the internal motion of the ^{15}N-$^1H^N$ bond vector using the following expression for the spectral density function:

$$J(\omega) = S^2\tau_m/(1+\omega^2\tau_m^2) + (1-S^2)\tau/(1+\omega^2\tau^2) \quad (19)$$

where ω is the ^{15}N Larmor frequency, S^2 is a general order parameter which describes the degree of spatial restrictions of motion, τ_m is the correlation time of the tumbling motion of the entire molecule ($1/\tau = 1/\tau_m + 1/\tau_e$) and τ_e is the effective correlation time for the reorientation of ^{15}N-$^1H^N$ vector due to internal motion.

The model has been refined by Clore et. al, (**52**, **53**) by introducing two distinct correlation times to describe the internal motion. Assuming that the correlation time describing the faster of the two time scales contributes a negligible amount to the relaxation, the modified spectral density function becomes:

$$J(\omega) = S^2\tau_m/(1+\omega^2\tau_m^2) + (S_f^2-S^2)\tau/(1+\omega^2\tau^2) \quad (20)$$

where S^2 is expressed as the product of two order parameters, $S_f^2S_s^2$, characterizing the fast and slow internal motion, respectively, and τ_s is the effective correlation time for the slow internal motion ($1/\tau = 1/\tau_m + 1/\tau_s$).

To model experimentally observed transverse relaxation rates, an additional term, R_{ex}, is often required to account for the contributions from processes other than dipole-dipole interactions and chemical shift anisotropy (i.e conformational exchange).

$$1/T_2 = 1/T_{2(DD)} + 1/T_{2(CSA)} + R_{ex} \quad (21)$$

where $T_{2(DD)}$ and $T_{2(CSA)}$ represent contributions from dipole-dipole and chemical shift anisotropy to transverse relaxation.

1.3.6.3 Inversion recovery – measurement of T_1

The pulse sequence used for T_1 measurements consists of two radio frequency pulses separated by a variable delay (Fig. 13).

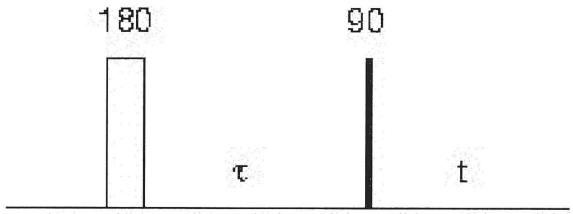


Fig. 13 Inversion-recovery pulse sequence

The first pulse in the sequence is a 180° pulse and generates –z magnetization (an inverted Boltzmann distribution). During the interval τ, magnetization relaxes back towards thermal equilibrium, its progress being monitored by a second 90° pulse which converts the population difference into coherence including the observable (-1) quantum coherence:

$$S(\tau, t) = a(\tau)\exp\{(i\Omega-\lambda)t\} \quad (22)$$

Where t is the acquisition time, Ω is the resonance offset and $a(\tau)$ is the amplitude of the NMR signal:

$$a(\tau) = a(\infty)(1-2\exp[-\tau/T_1] \quad (23)$$

which is negative for small values of τ and becomes positive for large values of τ (Fig. 14).

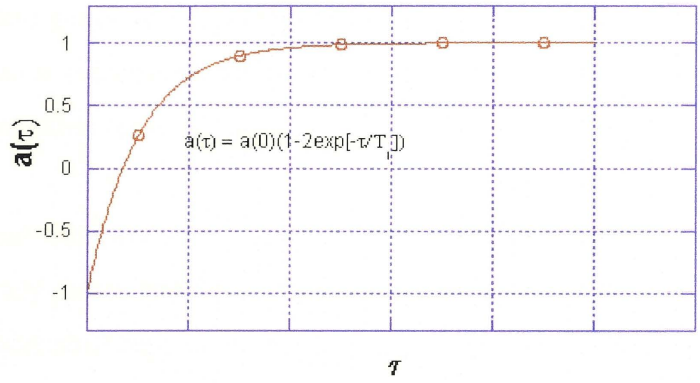


Fig. 14 Trajectory of the peak amplitude a as a function of τ in the inversion recovery experiment.

For the C-peptide, T_1 relaxation rates were measured in a 2D manner by placing a 180° pulse and a variable relaxation delay, τ, before the first 90° pulse of a TOCSY pulse sequence and were estimated for τ = 80 ms using the following formulas:

$$A(\tau) = A_o(1-2\exp[-\tau/T_1] \quad (24)$$

$$T_1 = \tau/(\ln 2+\ln(A/A_o)) \quad (25)$$

where $A(\tau)$ is the amplitude of the cross peak after a relaxation delay τ = 80 ms, and A_o is the absolute amplitude of the NMR signal measured after a relaxation delay τ = 1 ms.

1.3.6.3 Spin echo – measurement of T_2

The width at half height of a spectral peak (Hz) is given by $1/\pi T_2$, where T_2 is the transverse relaxation time. In the presence of field inhomogeneities, the width of NMR peaks are larger than this quantity. The spin-echo method allows the homogeneous decay due to homogeneous broadening to be distinguished from the inhomogeneous decay due to inhomogeneous broadening. The spin-echo experiment consists of the following element:

$$90_x^o - \tau - 180_x^o - \tau \text{ (FID)} \quad (26)$$

The initial 90_x° pulse generates the $-I_y$ operator from the equilibrium operator I_z. The overall effect of the spin echo segment '$\tau - 180_x^\circ - \tau$' is to take initial $-I_y/-S_y$ magnetization and generate I_y/S_y, magnetization so that evolution under the chemical shift Hamiltonian is refocussed. Therefore, evolution during the pulse sequence is due to the scalar interaction only.

In a heteronuclear spin echo experiment, the heteronuclear coupling can be refocussed by radio-frequency pulses applied to only one of the scalar coupled spins. In this way, both the chemical shift and heteronuclear couplings can be refocussed. When the experiment is performed for different values of τ a data matrix $S(\tau, t)$ is obtained. Fourier transformation with respect to t yields the spectrum $S(\tau, \Omega)$ where peak amplitudes vary as a function of τ as (Fig. 15):

$$a(\tau) = a(0)\exp[-\tau/T_2] \tag{27}$$

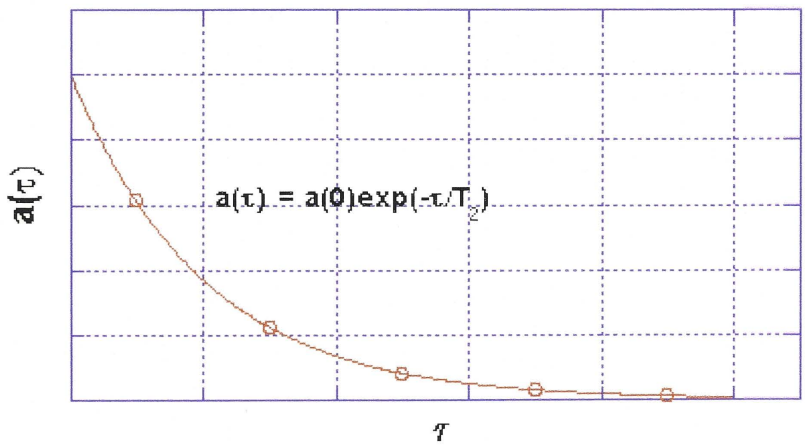


Fig. 15 Peak amplitude as a function of τ in a spin-echo experiment.

After a '$90_x^\circ - \tau - 180_x^\circ - \tau$' sequence all magnetization is aligned along the original axis. The '$\tau - 180_x^\circ - \tau$' sequence can be repeated multiple times until the signal disappears due to T_2 relaxation. An extreme case is presented by the mixing period of a TOCSY experiment where magnetization is transferred from one spin to another using a continuous train of 180° pulses with little spin precession between them.

In proteins, mobile residues exhibit enhanced flexibility resulting in long correlation times and long T_2 times. They can be easily identified in TOCSY spectra recorded with long mixing period because of their narrow resonances.

1.3.6.4 $^{15}N\{^1H\}$ steady-state NOE experiment

Overhauser (**54**) suggested an experiment where application of a weak radio frequency field at the Larmor frequency of one spin, for a sufficiently long time, has a strong effect on the longitudinal magnetization of the non-irradiated spins. This experiment is still exploited in heteronuclear (AX) spin systems for dynamics studies in the form of the steady-state nuclear Overhauser effect or steady-state NOE measurements.

For an ensemble of heteronuclear IS spin systems (assuming that relaxation is entirely dipolar in origin and a continuous radio frequency field is applied at the Larmor frequency of the I spin), with spin species I of gyromagnetic ratio γ_I coupled to a species S characterized by a gyromagnetic ratio γ_S, the NOE enhancement factor is given by:

$$\varepsilon_{NOE} = 1 + [(\gamma_I/\gamma_S)(W_2-W_0)/(W_0+2W_{1S}+W_2)] \quad (28)$$

where

$$W_{1S} = (3/20)b_{IS}^2 J(\omega_S) \quad (29)$$

$$W_2 = (3/5)b_{IS}^2 J(\omega_I+\omega_S) \quad (30)$$

$$W_0 = (1/10)b_{IS}^2 J(\omega_I-\omega_S) \quad (31)$$

are the single, zero and double quantum transition probabilities and

$$b_{IS} = -(\mu_0\gamma_S\gamma_I h /8\pi r_{IS}^3) \quad (32)$$

is the dipole-dipole coupling constant for the IS spin pair.

The NOE enhancement factor is a function of the correlation time (τ_c) and is often used to characterize the motion of molecules. For rapid molecular motions (small τ_c), the enhancement factor is given by:

$$\varepsilon_{NOE}(fast) = 1 + (\gamma_I/2\gamma_S) \quad (33)$$

It is equal to –3.93 for a $^{1}H^{N}$-^{15}N spin pair, showing that ^{15}N magnetization has a negative sign when the $^{1}H^{N}$ spins are saturated. For slow molecular motions characterized by large values of τ_c

$$\varepsilon_{NOE}(slow) = [(\gamma_I^2-\gamma_S^2)(3\gamma_I^2+5\gamma_I\gamma_S-10\gamma_S^2)/(3\gamma_I^4+\gamma_I^2\gamma_S^2-10\gamma_I\gamma_S^3+10\gamma_S^4)] \quad (34)$$

For a $^{1}H^{N}$-^{15}N spin pair, $\varepsilon_{NOE}(slow)$ is equal to 0.783 showing that ^{15}N magnetization is reduced by $^{1}H^{N}$ irradiation.

In order to characterize the backbone dynamics of proteins, $^{1}H^{N}$-^{15}N steady state NOE values are obtained by recording spectra with (NOE experiment) and without (NONOE experiment) the use of $^{1}H^{N}$ saturation applied before the start of the experiment. (Appendix I). The NOE is determined by

$$NOE = I_{sat}/I_{eq} \quad (35)$$

where I_{sat} and I_{eq} represent the measured intensities of a resonance in the presence and absence of proton saturation. This experiment can be used in a straightforward manner for the determination of folded and unfolded residues in a protein structure. In proteins, not all amino acids have the same effective correlation time due to local motions. Amino-acid residues in the rigid part of the protein backbone (characterized by longer correlation times) are identified by positive signals in the NOE experiment, while very flexible residues ($\tau_c \sim 10^{-9}$-10^{-10} s) yield negative signals. There is growing interest in flexible regions of proteins because they are sometimes involved in a diversity of functional roles, including intermolecular binding and recognition (**55**).

These regions can also be mobile in the crystalline state. In this case, they are not well resolved in the electron density maps used to determine the X-ray crystal structure. Thus, solution-phase NMR spectroscopy is the most powerful method to study flexible regions in proteins.

1.4 Secondary structure

1.4.1 Introduction

The secondary structure of a protein refers to the local conformation of its backbone and includes α-helices, β-sheets and turns. These secondary structural elements of proteins can be identified from the chemical shifts of the α protons, the relative

intensities of the NOE connectivities between $^1H^N$, $^1H^\alpha$ and $^1H^\beta$ protons and by three bond coupling information. α-helices are characterized by strong $d_{NN}(i, i+1)$ NOEs, weak $d_{\alpha N}(i, i+1)$, $d_{NN}(i, i+2)$ and the presence of $d_{\alpha N}(i, i+3)$ and $d_{\alpha\beta}(i, i+3)$ NOEs, while β-sheets elements are characterized by strong $d_{\alpha N}(i, i+1)$ and weak $d_{NN}(i, i+1)$ NOEs. Antiparallel β-sheets give long range $^1H^N(i)$-$^1H^N(j)$, $^1H^N(i)$-$^1H^\alpha(j)$ and $^1H^\alpha(i)$-$^1H^\alpha(j)$ cross peaks, where i and j represent residues from different β-strands. Parallel β-sheets are characterized by weak interstrand $^1H^N$-$^1H^\alpha$ and $^1H^N$-$^1H^N$ cross peaks.

Turn structures show medium range $d_{\alpha N}(i, i+2)$ and weak $d_{NN}(i, i+1)$ NOEs. Reverse turns or loops which reverse the direction of the polypeptide chain at the surfaces of molecules (i.e. β-hairpins and β-turns) are prevalent in globular proteins and are frequently classified as a third type of secondary structure. A sequential region of $^3J_{HNH\alpha} < 6$ Hz is indicative of α-helices, while regions of β-sheets have $^3J_{HNH\alpha} > 8$ Hz. The NOE patterns and scalar couplings observed for different types of secondary structure are summarized in Table 2.

NOE type	**α-helix**	**β-sheet**
$d_{NN}(i, i+1)$	strong	weak
$d_{\alpha N}(i, i+1)$	medium	strong
$d_{\beta N}(i, i+1)$	medium/strong	weak
$d_{\alpha N}(i, i+3)$	present	absent
$d_{\alpha\beta}(i, i+3)$	present	absent
$d_{\beta N}(i, i+3)$	present	absent
$^3J_{HNH\alpha}$	< 6 Hz	> 8 Hz
$^1H^\alpha$ shifts	Upfield of random coil shifts	Downfield of random coil shifts

Table 2 NOEs, $^1H^\alpha$ shifts and scalar couplings characteristic of secondary structure types.

The deviation of $^{1}H^{\alpha}$ chemical shifts from their random coil values can be used to support secondary structure information obtained from NOE connectivities since $^{1}H^{\alpha}$ chemical shifts are dependent on the nature of secondary structure (**56**).

In structured regions, the exchange of amide hydrogens with solvent is slowed due to hydrogen bonding. Protein interiors are largely excluded from contact with their surrounding solvent, and hence, amide proton exchange with solvent is slow for this type of amide proton. Exchange in these regions occurs via transient local unfolding, the rate of this exchange being a reflection of the conformational mobility in their surroundings. Hydrogen bond data can be incorporated as restraints in the later stages of structure refinement, yet only in well defined regions of secondary structure, where only one possible hydrogen-bond acceptor is consistent with the NOE data.

1.4.2 Chemical Shift Index (CSI)

Wishart et al. (1992) have demonstrated that the H^{α} proton experiences an upfield shift with respect to the random coil shift in the α-helix conformation, and a downfield shift in β-strand (**56**). If the H^{α} chemical shift for a specific residue in the protein is greater than the random coil H^{α} chemical shift by 0.1 ppm for that particular amino acid, it is assigned the label "1" and if it is less than the random coil H^{α} chemical shift by 0.1 ppm, it is labelled "-1". If the H^{α} chemical shift is within this range (for the random coil H^{α} chemical shifts), it is labelled "0". Any sequence of four or more "-1" that is not interrupted by "1" indicates an α–helical conformation and any sequence of three or more "1" that is not interrupted by "-1" indicates a β-strand conformation. All other regions are defined as random coil or unstructured.

1.4.3 Neighbouring residue effects on chemical shifts

Wang and Jardetzky (2002) have refined the above approach by including nearest neighbour residue effects, statistically determined from a large empirical chemical shift database (**57**). For an amino acid sequence XYZ, they defined two correction factors, $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ representing the effects on Y's chemical shifts from the preceding residue (X) and the following residue (Z), respectively, where X, Y, Z represent any of the 20 naturally occurring amino acids, n stands for $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$,

$^{13}C^{\beta}$, and $^{13}C'$ nuclei, and s represents the three secondary structural types: β-strand, α-helix and random coil.

The observed chemical shift of Y residue (δn,s) is composed of the following components:

$$\delta n,s = \delta n,coil + \Delta\delta n,s + \Delta(^{X}Y)n,s + \Delta(Y^{Z})n,s \quad (36)$$

where δn,coil is the random coil chemical shift of Y; Δδn,s is the secondary structural effect on Y's chemical shift, which is zero by definition when s = coil; $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ are the effects from the preceding residue (X) and the following residue (Z), where:

$$\Delta(^{X}Y)n,s = \Delta(^{X}Y)n,coil + \Delta\Delta(^{X}Y)n,s \quad (37)$$

$$\Delta(Y^{Z})n,s = \Delta(Y^{Z})n,coil + \Delta\Delta(Y^{Z})n,s \quad (38)$$

where $\Delta\Delta(^{X}Y)n,s$ and $\Delta\Delta(Y^{Z})n,s$ represent the variations of the nearest neighbour effects with secondary structure and they are zero by definition when s = coil. Thus the observed chemical shift (δn,s) can be expressed as:

$$\delta n,s = \delta n,coil + \Delta(^{X}Y)n,coil + \Delta(Y^{Z})n,coil + \Delta\Delta(^{X}Y)n,s + \Delta\Delta(Y^{Z})n,s \quad (39)$$

For each of the nuclei n, chemical shifts were categorized into three separate groups: β-strand, α-helix and random coil, based on the secondary structure of Y. $\Delta(^{X}Y)n,s$ for each of the three secondary structural types was calculated for the 20 naturally occurring amino acids X and Y as:

$$\Delta(^{X}Y)n,s = \langle\delta n,s(X)\rangle - \langle\delta n,s(w/o\ X)\rangle \quad (40)$$

where $\langle\delta n,s(X)\rangle$ and $\langle\delta n,s(w/o\ X)\rangle$ represent the average chemical shift of amino acid Y with and without amino acid X at the preceding position, respectively.
$\Delta(Y^{Z})n,s$ is determined similarly as:

$$\Delta(Y^{Z})n,s = \langle\delta n,s(Z)\rangle - \langle\delta n,s(w/o\ Z)\rangle \quad (41)$$

where $<\delta n,s(Z)>$ and $<\delta n,s(w/o\ Z)>$ represent the average chemical shifts of amino acid Y with and without amino acid Z at the following position, respectively.

The weighted averages of $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ over all 20 amino acids of Y, defined as $<\Delta(^{X}Y)n,s>$ and $<\Delta(Y^{Z})n,s>$ were calculated for each amino acid type of the neighbouring residue (X and Z) by:

$$<\Delta(^{X}Y)n,s> = \Sigma N(X)^{*}\ \Delta(^{X}Y)n,s/\Sigma N(X) \quad (42)$$

$$<\Delta(Y^{Z})n,s> = \Sigma N(Z)^{*}\ \Delta(Y^{Z})n,s//\Sigma N(Z) \quad (43)$$

where N(X) is the number of chemical shifts of Y preceded by X, and N(Z) is the number of chemical shifts of Y followed by Z found in the data base (**57**).

The average secondary structural effects, $<\delta n,s>$ were obtained by calculating the Y's chemical shift difference between the β-strand (or α-helix) and the random coil:

$$<\delta n,beta> = <\delta n,beta(corrected)> - <\delta n,coil(corrected)> \quad (44)$$

$$<\delta n,helix> = <\delta n,helix(corrected)> - <\delta n,coil(corrected)> \quad (45)$$

and the averaged variations of the nearest neighbour effects with secondary structure were obtained by calculating the differences:

$$<\Delta\Delta(^{X}Y)n,beta> = <\Delta(^{X}Y)n,beta> - <\Delta(^{X}Y)n,coil> \quad (46)$$

$$<\Delta\Delta(Y^{Z})n,helix> = <\Delta(Y^{Z})n,helix> - <\Delta(Y^{Z})n,coil> \quad (47)$$

The resulting increment system provides good estimates of random coil chemical shifts, talking into account their dependence on the amino acid sequence.

1.5 *Expression and ^{15}N labelling of proteins for ^{15}N NMR*

^{15}N NMR is a well-established and useful technique for structural and dynamic investigations of proteins especially due to the large chemical shift dispersion of ^{15}N resonances (**58, 59**). The low natural abundance and inherent insensitivity ($\gamma(^{15}N)/\gamma(^{1}H) = -0.101$) of this isotope have limited the use of ^{15}N NMR in the study of proteins. However, selective and uniform labelling techniques can be used to enrich by over 200 fold in this isotope. Bacteria grown on a minimal medium enriched with ^{15}N sources such as ammonia or glutamic acid will uniformly incorporate the isotope into the backbone and side-chain amide, amine, imidazole, indole, and guanidino groups of proteins (**60, 61**). Ammonia is the preferred nitrogen source for *E. coli* and the only inorganic form of this element on which the bacteria can grow. The backbone and side-chain functional groups in a protein often can be distinguished by their ^{15}N and ^{1}H chemical shifts but only rarely resonances can be specifically assigned on the basis of chemical shift alone (**62**). The assignment can, however, be simplified if only a single amino acid is labelled. Within this study, protein samples were synthesized with a cell-free *in vitro* protein expression system which can produce high yields of proteins per input mass of labelled amino acid without the need of auxotrophic bacterial strains (**65**). The cell-free protein expression produced selectively labelled protein in good yields, so that ^{15}N-HSQC spectra could be recorded directly without further purification and concentration (**39**). The *E. coli* peptidyl–prolyl cis-trans isomerase PpiB, expressed in this system with amino acid selective isotope labelling was analyzed to identify metabolites produced in side reactions of this expression system.

CHAPTER 2

Expression and ^{15}N labelling of proteins for ^{15}N NMR

2.1 Introduction

As mentioned previously, the backbone and side-chain functional groups in a protein often can be distinguished by their ^{15}N and ^{1}H chemical shifts but only rarely resonances can be specifically assigned on the basis of chemical shift alone (**62**). One solution is to label the protein at specific sites, except for the case of an amino acid which occurs only once in the protein sequence. The side-chains or peptide backbone of the protein can be selectively ^{15}N enriched by growing the host bacteria in a defined medium supplemented with one or more ^{15}N-labelled amino acids or amino acid precursors. A major drawback for the selective labelling method for proteins expressed *in vivo* is the time consuming purification of every selectively labelled sample and the higher cost of labelled amino acids versus the ammonium salts used for uniform labelling with ^{15}N. Various strategies to increase efficiency of such methods have focussed on recording ^{15}N-HSQC spectra of partially purified proteins or even *in situ* in whole *E. coli* cells (**63**, **64**). Adequate results have been obtained by the use of auxotrophic bacterial strains for expression (**66**). In cells, NMR spectra have broad signals due to the high intracellular viscosity, while cell lysis exposes the protein to temperature and chemical stress including exposure to cell debris and proteolytic enzymes. Within this study, an alternative method that uses a cell-free *in vitro* protein expression system to produce high yields of proteins per input mass of labelled amino acid without the need of auxotrophic bacterial strains is described (**65**). The cell-free protein expression produces selectively labelled protein in good yields and of high purity, so that ^{15}N-HSQC spectra can be recorded directly without further purification and concentration (**39**). The *E. coli* peptidyl–prolyl cis-trans isomerase PpiB was expressed in this system with amino acid selective isotope labelling to check the accuracy of the method by recording ^{15}N-HSQC experiment before and after sample ultrafiltration.

2.2 Materials and methods

The *E. coli* protein cytoplasmic peptidyl-prolyl cis-trans isomerase (PpiB) had been expressed by Dr Kiyoshy Ozawa and Dr Madeleine Headlam in a cell-free expression system following the protocol by Guignard et al. (**39**). For preparation of sample of ^{15}N-Cys, ^{15}N-Leu, ^{15}N-Phe, ^{15}N-Val and ^{15}N-Tyr, the labelled amino acids (1 mM each) replaced unlabelled L-Cys, L-Leu, L- Phe, L-Val and L-Tyr in all mixtures. NMR measurements were performed before and after ultrafiltration (4000 x g, 10 min) with buffer exchange (50 mM sodium phosphate, pH 6.8) and addition of 10% D_2O to provide a lock signal. All experiments were recorded at 25 °C on a Varian NMR spectrometer operating at 600 MHz ^{1}H frequency on samples containing 50 mM sodium phosphate in 90% H_2O/10% D_2O at pH 6.8.

2.3 Results and discussion

As the expression yields of PpiB were approximately 1 mg per ml of reaction volume, the concentration was sufficiently high to record ^{15}N-HSQC spectra directly with the reaction mixture. Ultracentrifugation and buffer exchange provided sufficient purification to render PpiB the only molecule in solution with ^{15}N labelled amides.

The new buffer contained 50 mM sodium phosphate at pH 6.8, corresponding to the sample conditions previously determined for the assignment of backbone resonances (**67**). Two major problems for the selective incorporation of ^{15}N isotope into a protein are isotopic dilution and incorporation of the ^{15}N label in undesired residues. Isotopic dilution results in a decrease of the ^{15}N enrichment of the labelled residue. It can arise from transaminase activity which exchanges the ^{15}N label with ^{14}N from other nitrogen-containing molecules. It can also arise from the presence of unlabelled amino acids in the reaction mixture. Transfer of the label into other molecules results from the metabolic conversion of ^{15}N-labelled amino acids. If the ^{15}N-label ends up in an amide bond, it gives rise to a signal which could be confused with the cross-peaks from the selectively labelled amino acid of the target protein. In *in vivo* expression systems these problems can be limited by repression of bacterial amino acid metabolism using defined media supplemented with essentially all amino acids and by use of bacterial hosts with mutations blocking the appropriate pathways of amino acid synthesis and degradation.

In the free-cell expression system, additional amide cross-peaks were invariably from compounds of low molecular weight which could be removed by ultrafiltration. This was demonstrated by ^{15}N HSQC experiments (recorded for about 7 to 14 hours) employing sweep widths of 5000 and 7000 (or 10000) Hz in the ^{15}N and ^{1}H dimensions. The number of extrapeaks observed before ultrafiltration are given in Table 3. Cross-peaks of residues 16, 107, 115, 150 and 155 could not be observed, presumably due to fast exchange between water and amide protons. No ultrafiltration was carried out for ^{15}N-Cys, ^{15}N-Leu and ^{15}N-Tyr labelled PpiB samples, since no extra peaks were observed in these spectra (Fig. 16).

Residue type (number of residues in the sequence)	Extra peaks	Missing peaks (residue number)
Phe (12)	2	16, 107
Val (16)	2	150, 155
Leu (5)	0	115
Cys (2)	0	-
Tyr (3)	0	-

Table 3. ^{15}N-HSQC data of spectra recorded on ^{15}N-Phe, ^{15}N-Val, ^{15}N-Leu, ^{15}N-Cys and ^{15}N-Tyr labelled PpiB samples.

Notably, no cross peaks were observed for excess ^{15}N-labelled amino acid, since the signals from α-amino groups are broadened beyond detection by exchange with water protons.

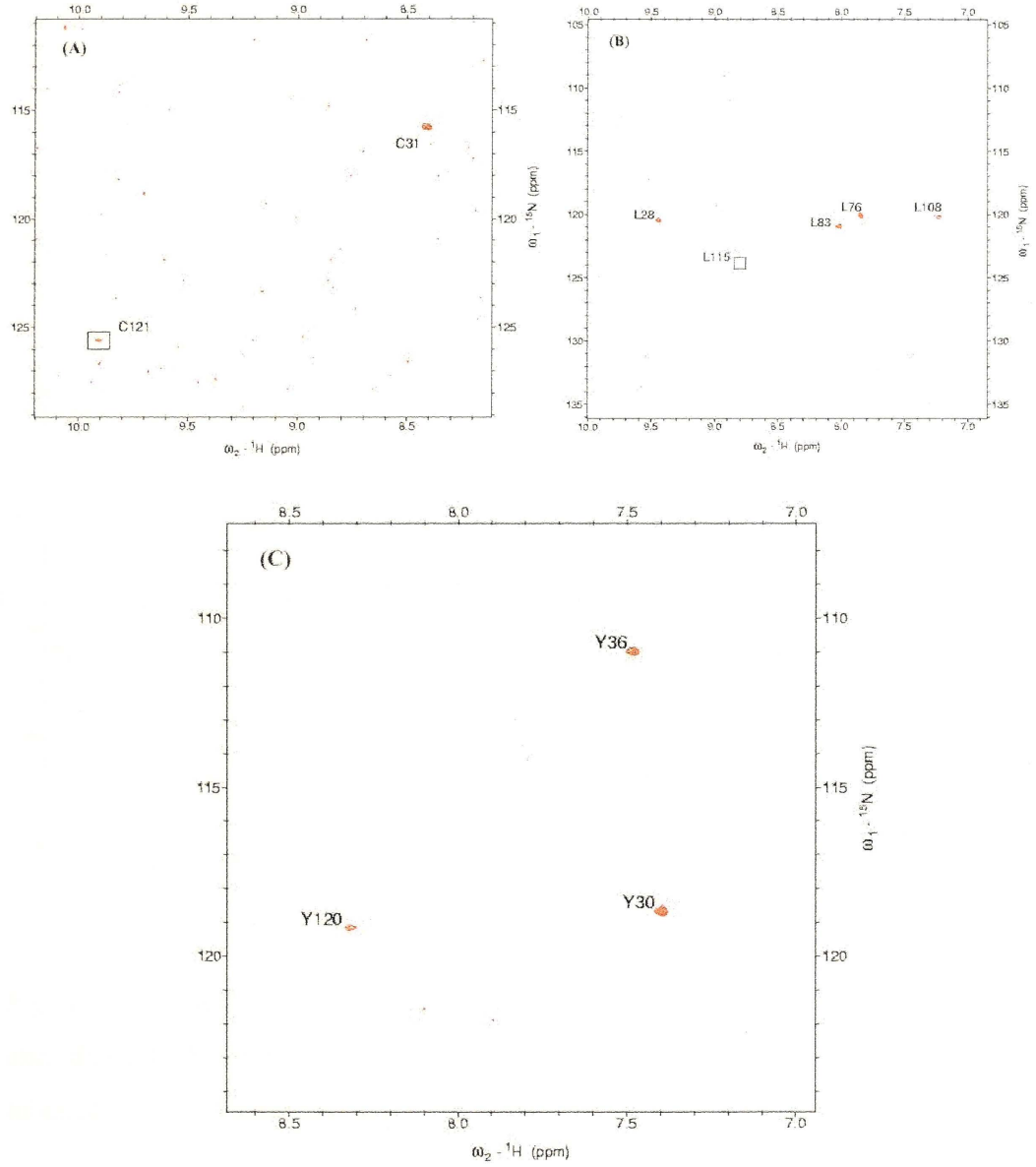


Fig. 16 ^{15}N-HSQC spectra of PpiB labelled with ^{15}N-Cys (A), ^{15}N-Leu (B) and ^{15}N-Tyr (C). Positions of unobservable residues are identified by boxes.

E. coli extracts contain enzymes that transform a fraction of some of the amino acids into low-molecular weight metabolites in which their α-amino groups are derivatized to amides, giving rise to extra peaks in the HSQC spectrum (Fig. 17).

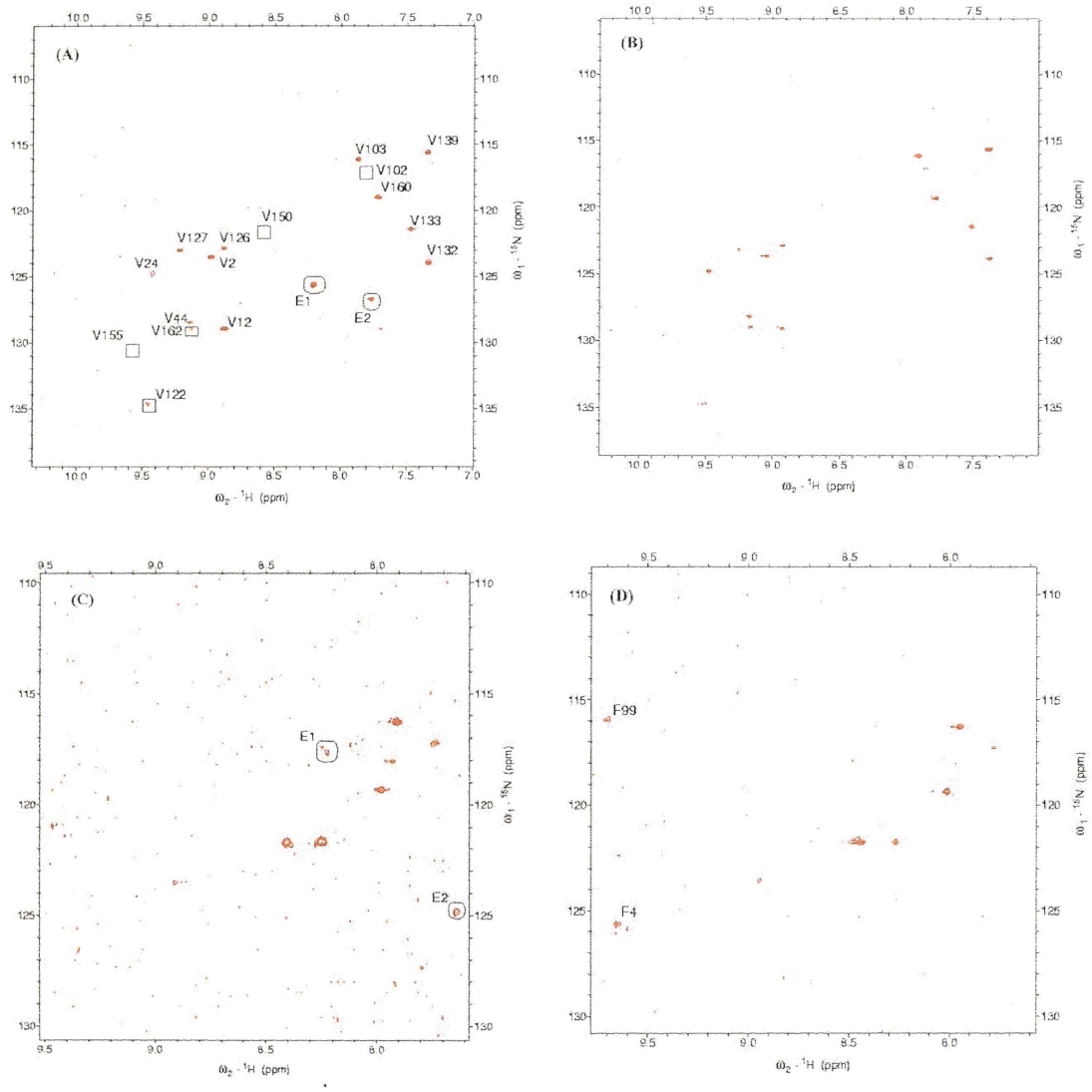


Fig. 17 ^{15}N HSQC spectra of PpiB labelled with ^{15}N-Val and ^{15}N-Phe before (A, C) and after (B, D) ultrafiltration and buffer exchange. Unexpected peaks are circled and labelled an E (extra peaks). The total recording time was approximately 14 h.

Some of the resonances (4 and 99 in Fig. 17D) could be observed after ultrafiltration combined with buffer exchange, while the signals arising from low-molecular weight compounds (circled resonances) were removed (Fig. 17B and 17D).

In conclusion, cell-free protein expression combined with ^{15}N NMR presents an inexpensive protein analysis method. In the case of PpiB, selectively labelled protein samples can be expressed in 0.5 ml of reaction medium using small quantities of

labelled amino acid and the samples can be analyzed by NMR without any further purification and with minimal sample handling.

CHAPTER 3

Structure analysis of proinsulin connecting peptide

3.1 Introduction

C-peptide is released from its precursors, preproinsulin and proinsulin, in the course of insulin biosynthesis (Fig. 18). Proinsulin is converted by a process of enzymatic cleavage to insulin and C-peptide in the beta cells. Two endopeptidases, prohormone convertases 2 and 3 (PC2 and PC3), cleave proinsulin at Lys^{64}-Arg^{65} and Arg^{31}-Arg^{32} sites, producing C-peptide and insulin (**68**). C-peptide was long thought to have no physiological effect on its own but recent data demonstrated that it binds with high affinity to cell surfaces, probably to G protein-coupled receptors, with subsequent activation of Ca^{2+}-dependent intracellular signalling pathways and stimulation of Na,K-ATPase activities (**1**). It was further shown that C-peptide is a biologically active peptide hormone with beneficial effects on kidney and nerve function in type 1 diabetes. (**2**).

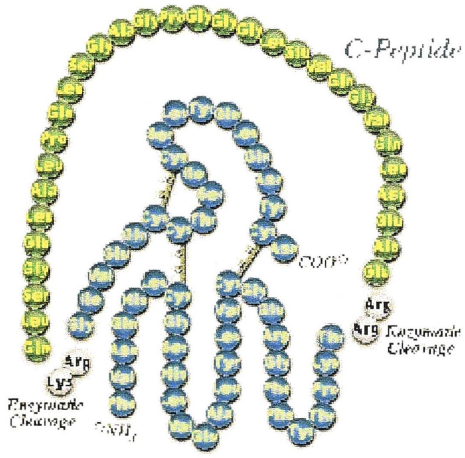


Fig. 18 Connecting peptide (C-peptide) is cleaved from proinsulin at the dipeptide bonds Arg-Arg and Lys-Arg to produce insulin. Two disulphide bonds connect insulin's A chain, composed of 21 amino acids, and B chain, composed of 30 amino acids (Diagnostic Products Corporation).

The ^{1}H resonances of the C-peptide have been assigned in TFE (trifluoroethanol) solution but not in water (**7**). Within this project, the ^{1}H, ^{13}C and ^{15}N NMR resonances of the C-peptide were assigned using a 36 mM sample in 90% H_2O/10% D_2O. Inter-residue NOEs were observed only between the neighbouring residues, showing that the C-peptide adopts a random coil conformation in solution (**69**). T_1 (^{1}H) relaxation rates also were in agreement with a random coil conformation. The observed chemical

shifts were compared to the chemical shifts predicted for random coil peptides (**8**). The differences between the observed and the predicted $^{1}H^{\alpha}$ chemical shifts were small, supporting the notion of a random coil conformation of the C-peptide in solution.

3.2 Materials and methods

100 mg of the C-peptide (provided by Prof. J. Wahren, Karolinska Institute – Sweden) was dissolved in 0.7 ml of 90% H_2O /10% D_2O at pH 6 to give a 36 mM solution. Inspection of the NMR spectra indicated no aggregation of the C-peptide at this concentration. All NMR experiments were recorded on a Varian NMR spectrometer operating at 600 MHz ^{1}H frequency. Spin systems were identified using two-dimensional (2D) double quantum filtered correlated spectroscopy (DQF-COSY) and total correlation spectroscopy (TOCSY). Sequence-specific assignments were completed using NOESY spectra. ^{15}N-HSQC, ^{13}C-HSQC, TOCSY-relayed ^{15}N-HSQC and TOCSY-relayed ^{13}C-HSQC spectra were used to assign ^{13}C and ^{15}N resonances and to resolve the overlap in the homonuclear spectra. Solvent suppression was achieved using the WATERGATE (water suppression by gradient tailored excitation) technique for TOCSY, WET for NOESY and presaturation for DQF-COSY. Typically, the recycle delay was 1 s and the number of transients was 8 (DQF-COSY) and 16 (NOESY, TOCSY and HSQC spectra). A 2D Varian pulse sequence designed for T_1 relaxation measurements was created by inserting a composite 180° pulse and a relaxation delay T at the beginning of a TOCSY pulse sequence. T_1 relaxation rates were calculated for a delay T = 80 ms according to:

$$M = (1-2e^{T/T1})*M_0 \qquad (48)$$

where M is the observed magnetization, and M_0 is the equilibrium magnetization.

Spectra were acquired with 2048 (DQF-COSY) in the F_2 and 1024 (NOESY, TOCSY and HSQCs) complex points in the F_1 dimension. The DIPSI scheme was used for mixing (100 ms) in TOCSY. The 2D data were processed and analyzed using NMRPIPE (**70**) and SPARKY (**71**) software packages.

3.3 *Results and discussion*

Sequential assignments were made according to the methods described by Wüthrich (1986) and are depicted in Figs. 19 and 20. The chemical shifts of ^{1}H, ^{13}C and ^{15}N are summarized in Table 4.

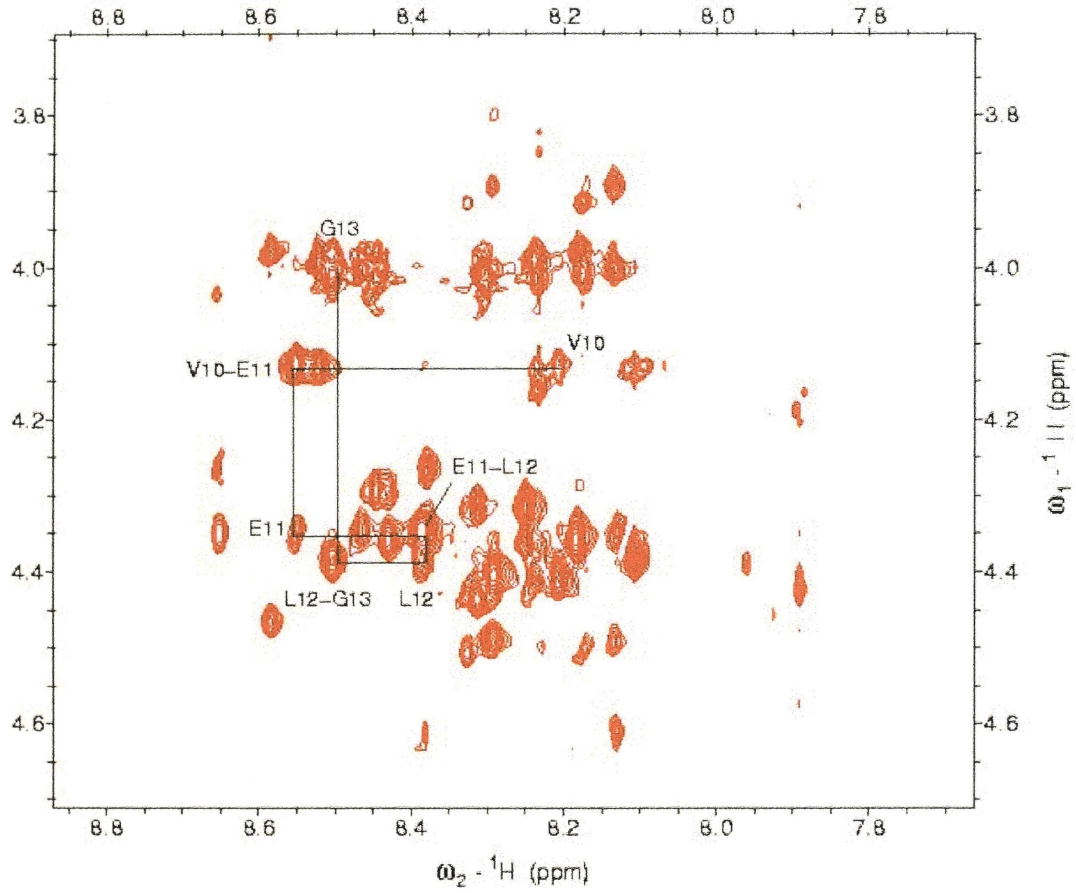


Fig. 19 ^{1}H NOESY spectrum of C-peptide recorded at 25 °C in water; the region ω_2 = 7.5 – 8.9 ppm and ω_1 = 3.6 – 4.7 ppm is displayed, showing sequential connectivities for the peptide fragment V10-G13.

2D heteronuclear spectra (^{15}N-HSQC, ^{13}C-HSQC, TOCSY-relayed ^{15}N-HSQC and TOCSY-relayed ^{13}C-HSQC) were used to assign ^{13}C and ^{15}N resonances and to resolve the overlap in the homonuclear spectra. In the ^{15}N-HSQC the residues Ala2 and Ala25 showed the same ^{15}N and $^{1}H^{N}$ chemical shifts. The $^{1}H^{\alpha}$ chemical shifts, however, were slightly different and could be identified in the DQF-COSY spectrum.

Amino acid residues Gly8, Gly13, Gly14, Gly15, Gly17, Gly19, Gly28, Ser20 and Ser29 were easily identified due to their characteristically upfield ^{15}N chemical shift (~100 ppm for Gly and 110 for Ser). Val7, Val10, Leu5, Leu12, Leu21, Leu24, Leu26 and Leu31 were recognized in the TOCSY spectrum by their two methyl resonances which yielded a characteristic row of double signals between 0 and 1.5 ppm. The Glu and Gln residues could only be distinguished through the sequential assignment process.

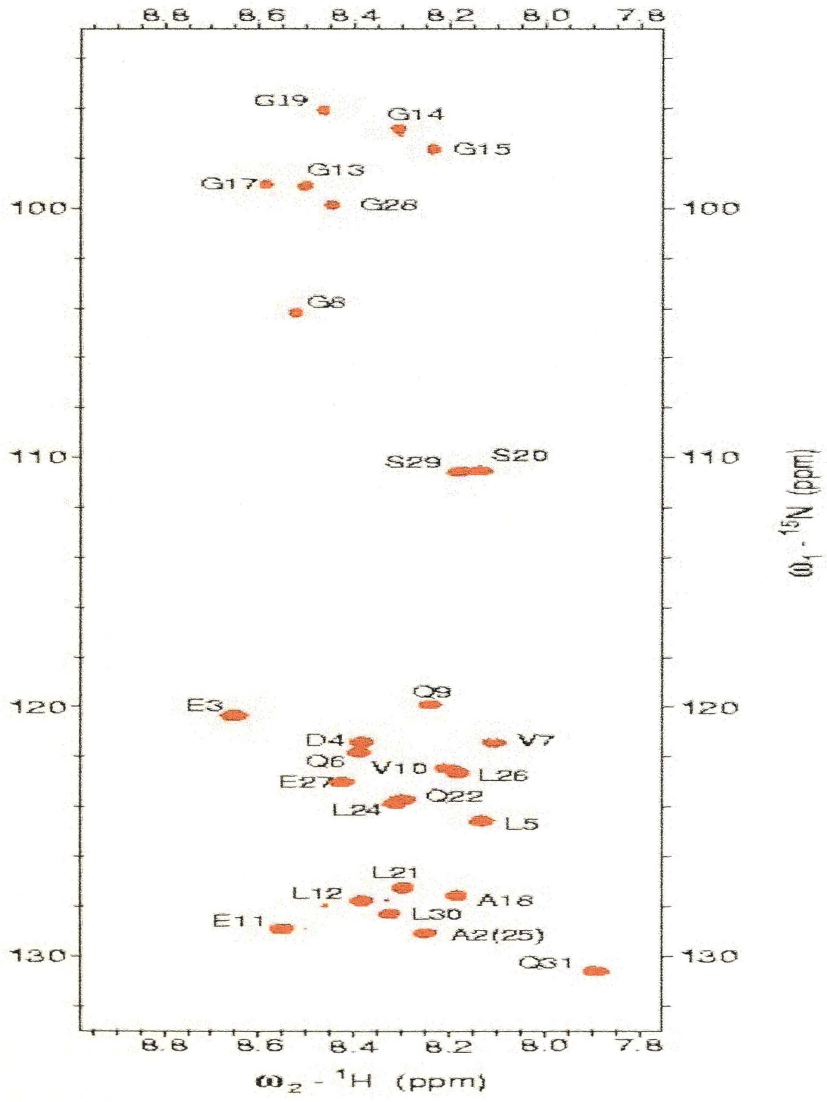


Fig. 20 ^{15}N-HSQC spectrum of C-peptide recorded at 25 °C in water; the region ω_2 = 7.7 – 9.2 ppm and ω_1 = 93 – 133 ppm is shown. The TOCSY-relayed ^{15}N-HSQC spectrum of C-peptide is shown in Fig. 8.

Table 4 Chemical shifts of C-peptide in 90%H_2O/10%D_2O at 25 ^{0}C

	CA	CB	CD	CD1	CD2	CG	HA	HB	HB1	HB2	HD1	HD2	HG	HG1	HG2	HN	N
Glu1	55.98	30.39	-	-	-	35.91	4.093	-	2.202	2.127	-	-	-	2.456	2.435	-	-
Ala2	52.47	19.22	-	-	-	-	4.370	1.413	-	-	-	-	-	-	-	8.249	129.0
Glu3	57.19	30.19	-	-	-	36.33	4.272	-	2.067	1.984	-	-	-	2.316	4.427	8.652	120.4
Asp4	54.50	41.11	-	-	-	-	4.619	-	2.753	2.652	-	-	-	-	-	8.382	121.4
Leu5	55.46	42.52	-	25.10	23.57	27.04	4.355	-	1.699	1.682	0.9565	0.8914	1.632	-	-	8.132	124.6
Gln6	55.98	29.32	-	-	-	33.91	4.392	-	2.144	2.038	-	-	-	2.400	2.389	8.390	121.8
Val7	62.93	32.79	-	-	-	21.20	4.142	2.130	-	-	-	-	0.9764	-	-	8.107	121.4
Gly8	45.51	-	-	-	-	-	4.002	-	-	-	-	-	-	-	-	8.522	104.2
Gln9	55.98	29.68	-	-	-	33.96	4.419	-	2.132	2.031	-	-	-	2.374	2.374	8.243	119.9
Val10	62.65	33.00	-	-	-	20.62	4.141	2.092	-	-	-	-	0.9771	-	-	8.207	122.5
Glu11	56.55	30.24	-	-	-	36.17	4.357	-	2.057	1.979	-	-	-	2.286	2.286	8.550	128.9
Leu12	55.46	42.52	-	25.10	23.57	27.05	4.395	-	1.722	1.694	0.9575	0.9005	1.645	-	-	8.384	127.8
Gly13	45.54	-	-	-	-	-	4.016	-	-	-	-	-	-	-	-	8.506	99.11
Gly14	45.29	-	-	-	-	-	4.029	-	-	-	-	-	-	-	-	8.306	96.81
Gly15	44.72	-	-	-	-	-	4.166	-	-	-	-	-	-	-	-	8.234	97.65
Pro16	63.79	32.13	49.92	-	-	27.28	4.469	-	2.310	2.025	3.689	3.663	-	-	-	-	-
Gly17	45.45	-	-	-	-	-	3.983	-	-	-	-	-	-	-	-	8.587	99.05
Ala18	53.10	19.26	-	-	-	-	4.129	1.442	-	-	-	-	-	-	-	8.184	127.5
Gly19	45.59	-	-	-	-	-	4.012	-	-	-	-	-	-	-	-	8.383	96.09
Ser20	58.52	64.16	-	-	-	-	4.502	3.901	-	-	-	-	-	-	-	8.137	110.5
Leu21	55.41	42.47	-	25.11	23.56	27.04	4.415	-	1.688	1.674	0.9533	0.8989	1.625	-	-	8.295	127.3
Gln22	57.45	29.06	-	-	-	33.58	4.656	-	2.126	1.975	-	-	-	2.412	2.412	8.292	123.7
Pro23	63.46	32.17	50.77	-	-	27.51	4.441	-	2.331	1.943	3.799	3.706	-	-	-	-	-
Leu24	55.51	42.44	-	25.11	23.57	27.05	4.322	-	1.679	1.679	0.9694	0.9208	1.604	-	-	8.315	123.9
Ala25	52.47	19.22	-	-	-	-	4.368	-	-	-	-	-	-	-	-	8.250	129.0
Leu26	55.35	42.52	-	25.10	23.57	27.05	4.338	-	1.669	1.669	0.9654	0.9104	1.626	-	-	8.181	122.6
Glu27	57.16	30.16	-	-	-	36.21	4.300	-	2.094	2.026	-	-	-	2.320	2.320	8.405	123.0
Gly28	45.59	-	-	-	-	-	4.030	-	-	-	-	-	-	-	-	8.446	99.88
Ser29	58.55	64.20	-	-	-	-	4.513	3.921	-	-	-	-	-	-	-	8.176	110.6
Leu30	55.52	42.44	-	25.11	23.57	27.05	4.434	-	1.719	1.685	0.9660	0.9086	1.668	-	-	8.326	128.3
Gln31	57.45	30.63	-	-	-	34.41	4.196	-	2.162	1.960	-	-	-	2.319	2.319	7.891	130.6

The analysis of $^{1}H^{\alpha}$ chemical shifts by the CSI (chemical shift index) method and comparison with the predicted $^{1}H^{\alpha}$ proton chemical shifts for random coil peptides indicated that C-peptide adopts a random coil conformation under these conditions (Fig. 21). A C program was written to assist the chemical shift prediction for random coil peptides, using the parameters reported by Wang et al. (**8**). The program requires the amino acid sequence of the peptide as an input file, and provides the predicted chemical shifts in an output file (see Appendix II).

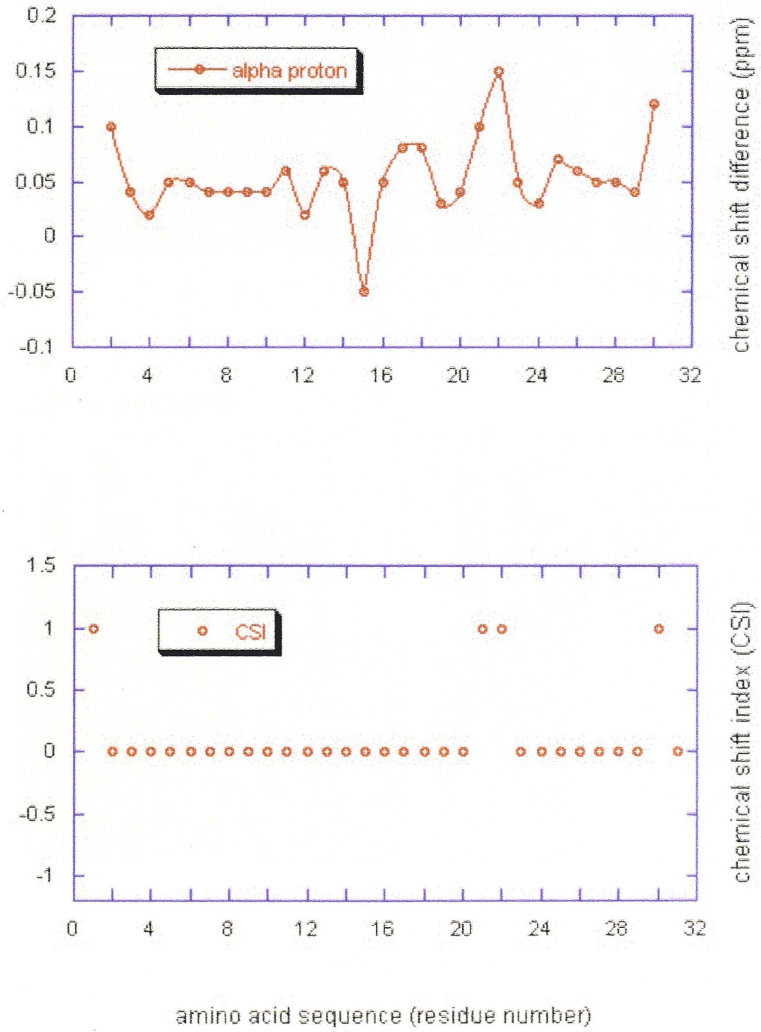


Fig. 21 Deviation of C-peptide H^{α} chemical shifts in water from the predicted random coil values and the resulting CSI values.

$T_1(^1H)$ relaxation rates were measured using delays of 1, 40 and 80 ms, and were estimated for a relaxation delay T = 80 ms according to equation (48). T_1 relaxation times for $^1H^{\alpha}$ and $^1H^N$ protons were almost uniform over the molecule (Fig. 22).

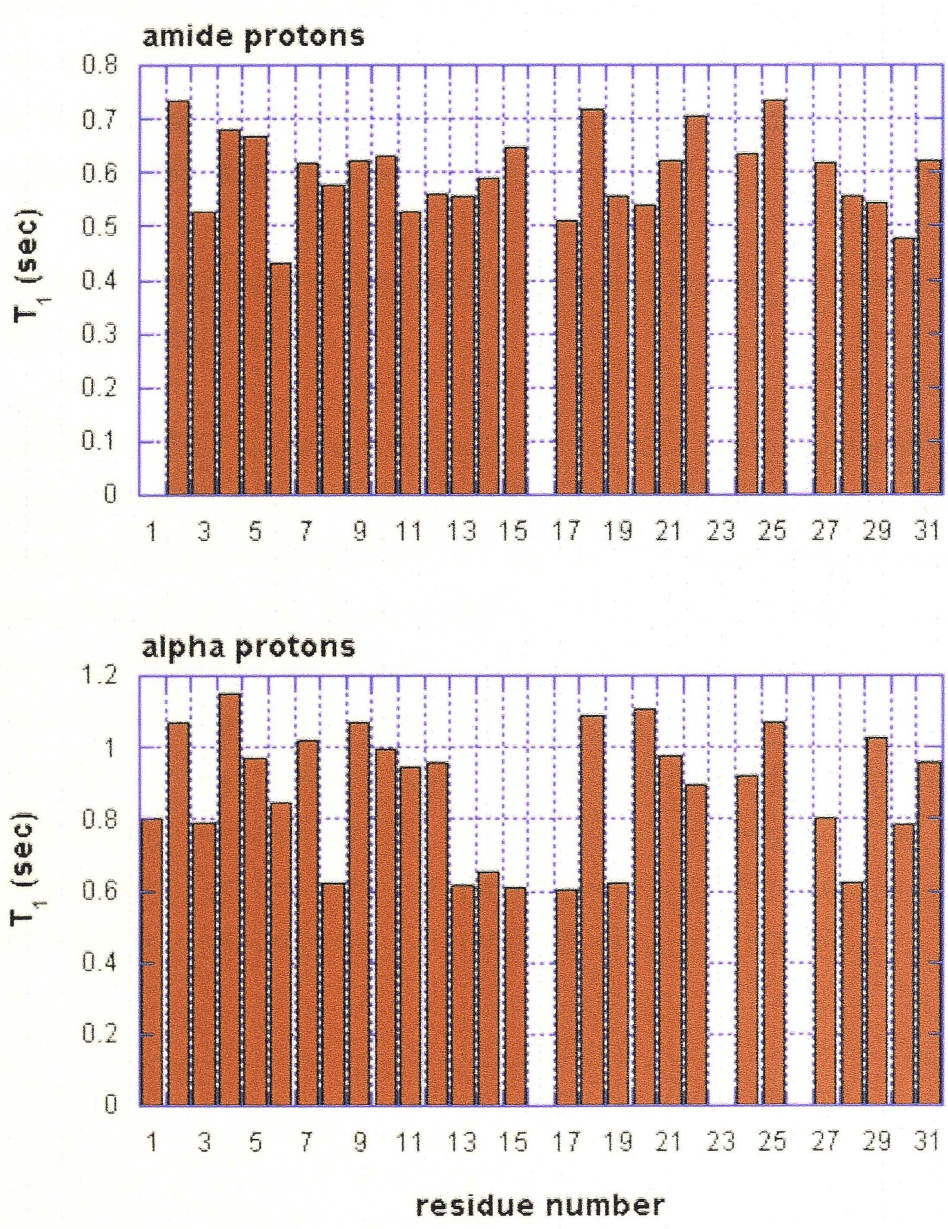


Fig. 22 Measured T_1 relaxation times for $^1H^N$ and $^1H^{\alpha}$ protons of C-peptide in water at pH 6 and 25 °C. A table with the measured T_1 relaxation times for all protons is provided in Appendix II.

The intra-residue NOEs in combination with relaxation measurements, chemical shift index and chemical shift prediction methods point to a random coil structure of the C-

peptide in solution. The C program developed within this project might be a useful tool for the assignment of random coil peptides (see Appendix II). The output is written as a list of chemical shifts in Sparky file format and can be loaded directly onto a spectrum. The results are in agreement with previous studies of the C-peptide; CD data showed that C-peptide exhibits a random coil structure in water and micelles, while the analysis of C-peptide H^{α} chemical shifts in 95% TFE/5% H_2O showed that the first 11 amino acid residues form a helical structure and residues 12-31 are in a flexibly disordered random coil conformation (**7**).

Specific binding of C-peptide to cell membranes has been studied using fluorescence correlation spectroscopy (**72**).

Addition of unlabelled C-terminal pentapeptide, which is known to possess more than 90% of the molecule's ability to stimulate Na^+,K^+-ATPase (**1**), was accompanied by displacement of bound rhodamine-labelled C-peptide, with similar effects as a control experiment that utilized the intact C-peptide. This indicates that the C-terminal fragment is involved in the binding process (**72**). Furthermore, exposure of the cells to pertussis toxin, which is known to affect the cysteine residue in the C-terminal chain of the α subunit of G protein (**73**), affects the stimulation of C-peptide on Na^+,K^+-ATPase providing evidence for the involvement of G protein in the signal transduction pathway. In conclusion, the apparent specificity of receptor recognition does not depend on a defined conformation of the free peptide in solution.

CHAPTER 4

Structure analysis of the C-terminal domain of E. coli DnaG primase

4.1 DnaG-C

The DnaG primase is a single stranded DNA-dependent RNA polymerase and is an essential component of the replisome (Fig. 23), the machine that accomplishes chromosomal DNA replication in *Escherichia coli* (**10**). It is the enzyme responsible for repeated synthesis of short oligoribonucleotide primers that are extended by DNA polymerase III holoenzyme to produce Okazaki fragments on the lagging strand at the replication fork, and it also synthesizes the initiating RNA primers on both leading strands at the origin of replication (*ori*C). It is composed of three domains, a small N-terminal zinc binding domain, a larger central domain responsible for RNA synthesis comprising residues 111-433, and a C-terminal domain comprising residues 434 – 581 (DnaG-C, Fig. 24) that is believed to interact with the hexameric DnaB helicase (**11**).

The site of DnaG interaction in *E. coli* DnaB has not been precisely defined, but it appears to involve the N-terminal domain of the helicase, whose structure has been determined independently by X-ray crystallography (**83**) and NMR (**84**). ^{15}N-HSQC experiments were recorded on a ^{15}N-labelled DnaB-N sample with and without unlabelled DnaG-C present in solution (Guido Pintacuda, unpublished results). No significant changes in the position of the observed resonances was observed, indicating the absence of an interaction between these two domains.

DnaG-C was expressed in *E. coli* under the control of tandem bacteriophage λ promoters and the protein was purified in yields of 4-6 mg/L of culture and studied by X-ray crystallography and NMR (**12**). The crystal structure revealed the presence of a small globular domain at the N-terminus and a helix hairpin at the C-terminus. A long helical segment between both subdomains is highly solvent exposed, suggesting that this segment can be flexible in solution. This notion is further supported by the observation of two slightly different structures in the unit cell where the interconnecting helical segment is bent to different degrees.

Previously published studies on calmodulin, a protein composed of two globular domains connected by a long flexible linker, showed a long continuous helical conformation for this interconnecting segment in crystals and a flexible structure in solution (**85, 86, 87, 88**). In crystals, however, it was also shown that the interdomain

linker loses its helical character upon binding to target peptides, with the two domains coming closer and clamping the recognition peptide between their exposed hydrophobic cores (**86**). Considering the flexibility of the interdomain linker of calmodulin in solution, it seems well possible that a similar situation prevails in DnaG-C. DnaG-C is expected to be monomeric in solution since the full-length primase is monomeric under all conditions that have been examined (**13**). In the crystal structure, however, two molecules are closely intertwined in an approximately antiparallel manner. The present study was performed to identify flexible peptide segments of DnaG-C in solution and to assess the structural integrity of the interconnecting helix.

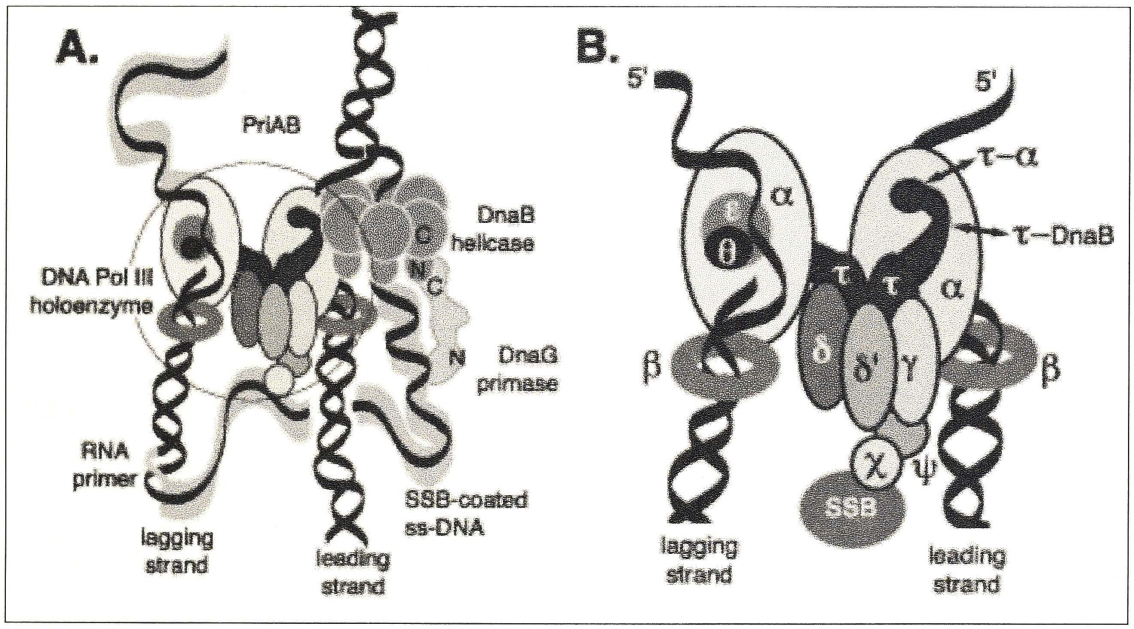
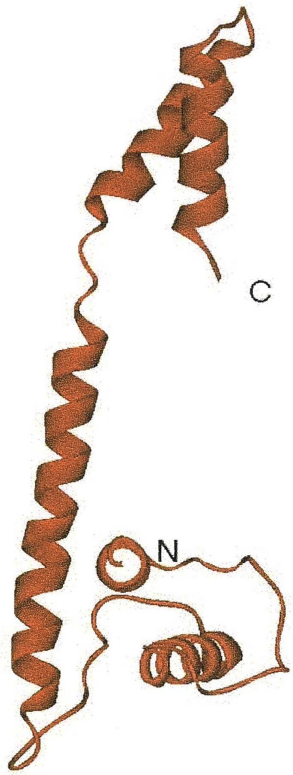


Fig. 23 Schematic representation of DNA replisome (**74**) (A) and DNA polymerase III holoenzyme (B). The primosome includes the DnaB helicase, PriA and PriB proteins, and the DnaG primase (**80**, **81**, **82**), the polymerase III holoenzyme (Pol III), comprising the core subunit α responsible for DNA synthesis, ε subunit (proofreading exonuclease), β and θ subunits (processivity-determining sliding clamp) and τ subunit required for dimerization of Pol III (**75**, **76**, **13**, **77**, **78**, **79**) (B) and other proteins like single stranded DNA-binding protein (SSB) required to ensure the processivity of Pol III and to support helicase activity, DNA gyrase which acts as a swivel to relieve topological tension in advance of helicase action, DNA ligase and DNA polymerase I to ensure contiguity of the lagging strand (**10**). (Figure provided by Dr N. E. Dixon)



1 10 20 30 40
AAESGVSR*PV*PQLKRT**{TMRILIGLLVQN*P*ELATLV}***PP***LENLDENK
50 60 70 80
L*P*GL**{GLFRELVNTCLS}**Q*P*GLTTGQLL**{EHY}**RGTNNA**{ATLEKLSMW**
90.....100 110 120 130
DDIADKNIAEQTFTDSLNH}MFD**{SLLELRQEELIARER}**THGLSN**{EE**
140
RLELWTLNQEL}A**KK**
148

Fig. 24 X-ray crystal structure and the amino acid sequence of DnaG-C consisting of helices comprising residues: 17-36, 50-61, 72-74, 81-108, 112-126, 132-145. Flexible regions and alpha helices are illustrated in red and blue respectively. (Figure provided by Karin Loscha).

4.2 Materials and Methods

4.2.1 Overproduction and purification

The protein was expressed and purified by Karin Loscha following a protocol previously described by Dr Dixon's group (**12**).

4.2.2 NMR measurements

All experiments were recorded in 90%H_2O/10% D_2O at 36 °C using a Varian NMR spectrometer operating at 600 MHz ^1H frequency. NOESY experiments were recorded on a 0.5 mM sample in 10 mM phosphate at pH 6 and 7 using 1024 complex points in the t_2 domain and 512 increments in t_1. The TOCSY experiment (80 ms mixing time, spectral width (sw) = 10000 Hz, number of points (np) = 1024, number of transients (nt) = 48, and number of increments (ni) = 200) was recorded at pH 6 and 25 ^{0} C using a sample of 2.26 mM unlabelled protein. ^{15}N-HSQC (np = 1024, nt = 16, ni = 128) and $\{^1H\}^{15}N$ NOE (2.9 s presaturation delay) experiments employing sweep widths of 8000 Hz and 3600 Hz in ^1H and ^{15}N dimension, respectively, were acquired on a $^{13}C/^{15}N$–labelled sample at pH 6, in 10 mM phosphate.

For the backbone resonance assignment a standard set of 3D triple-resonance experiments including: HNCA[4400, 1800, 60x40], HNCO[2100, 1800, 48x44], CBCA(CO)NH[12070, 1800, 68x44], NOESY-^{15}N-HSQC [8000, 1800, 48x44], HNCACB[10070, 1800, 50x40], and HNCAHA[3000, 1800, 80x44], with [sw_1, sw_2, t_1xt_2], were recorded. The number of scans was: 8 for HNCA, HNCO, CBCA(CO)NH, HNCACB, HNCAHA and 16 for TOCSY-^{15}N-HSQC. The total experimental time was approximately two weeks.

4.2.3 Spectral Processing

Data processing and analysis was performed using PROSA 2.8 (**89**) and XEASY 1.4 (**90**) software packages. NMR spectra were referenced using the water resonance.

4.2.4 *Spectral Analysis*

Spectral analysis and peak picking were performed using the XEASY software. Each cross peak in the 2D ^{15}N-HSQC spectrum was labelled and the resulting peak list was loaded into the 3D spectra. A strip was defined for each cross peak and searched for additional overlapping peaks that could not be identified in the ^{15}N-HSQC spectrum. Cross peaks resulting from scalar coupling connectivities were assigned to the side-chain atoms of the spin system. Spectral correlations between strips were determined starting with a strip that corresponded to a known amino acid spin system (i.e. Ala and Gly residues were easily identified in the HNCA and CBCA(CO)NH spectra and were starting points for the sequential assignment process). The strips were printed and inspected visually to identify the most likely sequential neighbour. When deciding on the most probable sequential neighbour, detailed comparison of the line shapes of the intra-residual cross peaks and the potential sequential cross peak were made. This procedure was repeated until almost the entire sequence of spin systems could be mapped onto the amino acid sequence.

4.3 *Results and discussion*

NMR experiments with several samples of unlabelled DnaG-C were recorded to investigate the tertiary structure of the protein, to identify mobile residues at its termini and to establish the necessity of further experiments with isotopically labelled samples. A 2D TOCSY (80 ms mixing time) spectrum recorded at 25 °C on a 2.26 mM sample, pH 7, identified 12 mobile residues at the C- and N-termini (Fig. 25) suggesting that the structured core of the protein comprises residues 11-146 (residues 444-579 of DnaG).

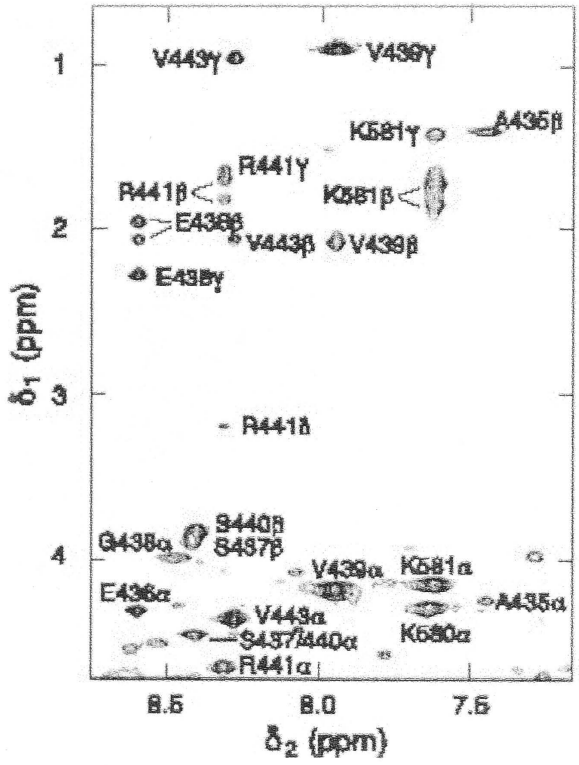


Fig. 25 Spectral region from the 2D TOCSY NMR spectrum recorded with 2.26 mM DnaG-C at pH 7 and 25 °C. Cross-peaks are labelled using the one letter code followed by the position in the DnaG sequence.

A NOESY spectrum of DnaG-C at 36 °C on a 0.5 mM sample in 10 mM phosphate, pH 7, showed broad lines for most of the cross peaks, indicating aggregation of the protein under these conditions (data not shown). Spectra recorded using unlabelled and $^{13}C/^{15}N$-labelled samples at pH 6 in the presence of 100 mM NaCl showed narrower linewidths characteristic of a largely monomeric protein.

Consequently, all experiments (^{15}N HSQC and a set of 3D triple-resonance experiments, described above) were recorded at pH 6 and 36 °C in buffer containing 10 mM phosphate and 100 mM NaCl. The ^{15}N-HSQC spectrum of DnaG-C is shown in Fig. 26 with 136 (the amide proton signals of residues 2, 14, 16, 131 and 132 were not observable, presumably due to exchange broadening) of the 141 expected backbone proton amide resonances identified. The spectrum exhibits a region between 7.6 ppm and 8.7 ppm that contains some severely overlapping resonances.

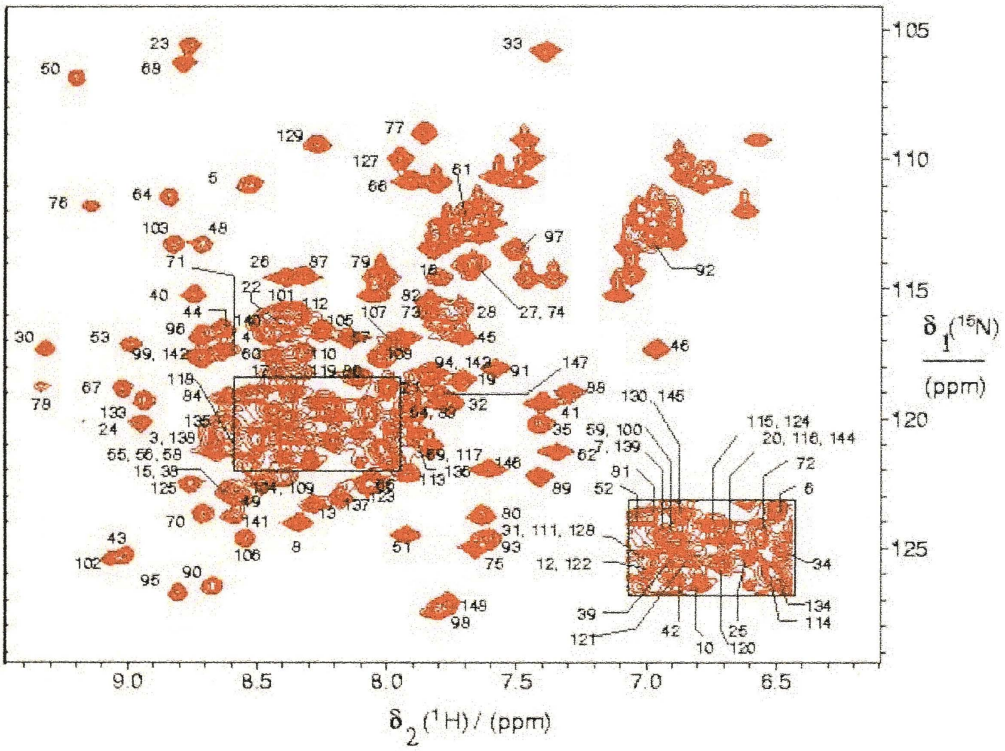


Fig. 26 The assigned ^{15}N HSQC spectrum recorded on 0.6 mM ^{13}C/^{15}N DnaG-C at pH 6 and 36 °C. Cross peaks are labelled according to the position in the sequence starting from 1 to 148. A table with chemical shifts and the amino acid sequence is provided in Appendix III.

Ala and Gly residues were easily identified in CBCA(CO)HN (**42**), HNCA (**35**) and HNCACB (**91**) spectra and were used as the starting points for sequential assignment.

The $^{1}H^{N}$ and ^{15}N resonances of residue (i) were correlated with $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ spins of the preceding residue (i-1) using the CBCA(CO)HN experiment and matched with the

intra- and inter-residues $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances observed in the HNCA and HNCACB data. Further, a 100 ms mixing time NOESY-^{15}N -HSQC (**92**) spectrum provided $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$, confirming the sequential assignment made from triple resonance data (Fig. 6). The overlap of residues: 55-56-57, 15-38, 115-124, 20-116-144, 130-145, 111-128 and 12-122 was resolved using the HNCO (**43**) spectrum.

The secondary structure obtained by short and medium range NOE contacts was compatible with the X-ray structure (Fig. 24). Less than 10% of the protein precipitated during a week in the spectrometer.

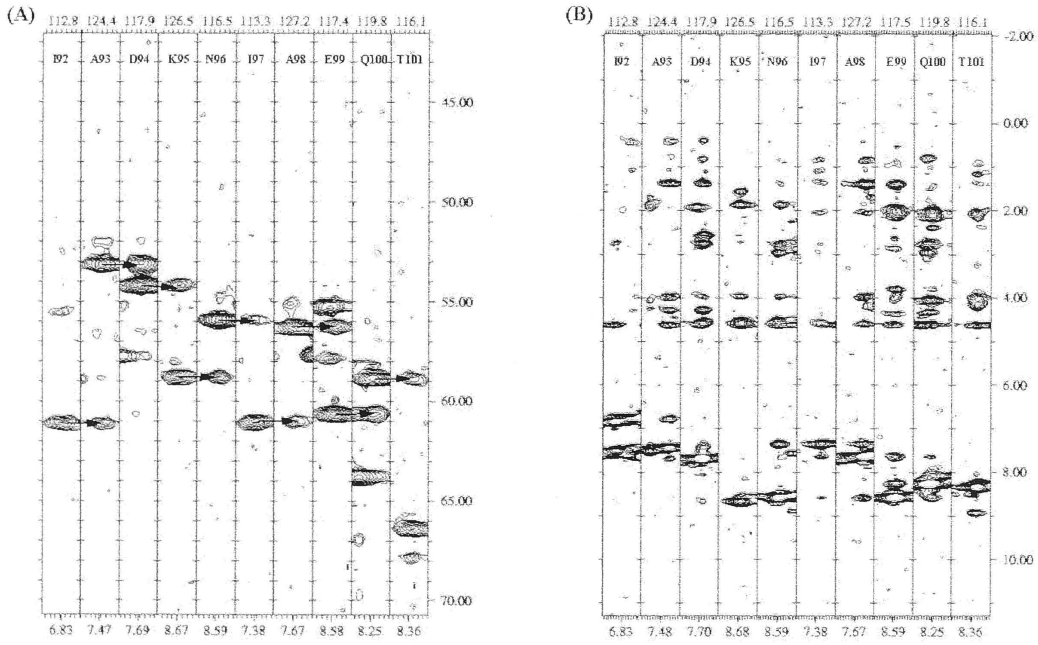


Fig. 27 F_1 ($^{13}C^{\alpha}$ and ^{1}H) strip plots of the HNCA (A) and NOESY-^{15}N-HSQC (B) spectra of DnaG-C showing sequential connectivities obtained for the peptide fragment I92 – T101. The sequential connectivities derived from the analysis of the HNCA spectrum is indicated by arrows connecting the corresponding intra- and inter-residue cross peaks.

The steady-state NOEs were obtained by recording modified HSQC spectra with and without a proton presaturation delay (2.9 s) (**93**) and were determined from the ratio of the peak intensities with (I_{sat}) and without (I_{eq}) proton presaturation (Fig 28). Average values of NOE ~ 0.7 were found for residues of the folded core (10–146).

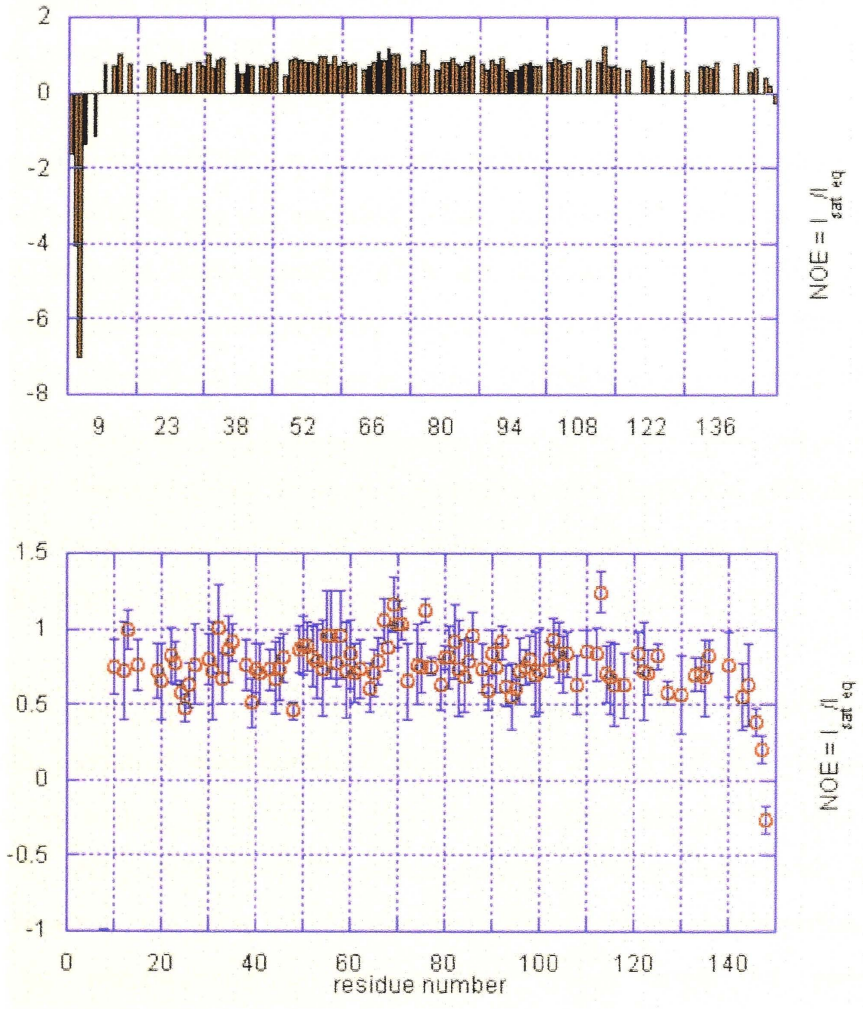


Fig 28 $^{1}H\{^{15}N\}$ NOE values were calculated for 115 residues of DnaG-C from the ratio of the peak intensities recorded with (I_{sat}) and without (I_{eq}) proton presaturation. ^{15}N-HSQC experiments, with and without 2.9 s proton saturation before the first 90^{0} ^{1}H pulse were recorded at 36 °C. Errors (bottom) were estimated from the background noise signal. Peak intensities were measured using the peak-picking subroutine in Sparky. A table with the measured NOEs is provided in Appendix III.

These data confirmed a rigid structured protein for the assigned peptide fragments. In particular the NOE data indicate the absence of increased mobility for the helix connecting the two subdomains of DnaG-C on a nanosecond timescale. Negative heteronuclear NOEs were obtained for the mobile residues at the N- and C- termini (peptide fragments A1-V10, K147-K148), in agreement with the small number of resonances observed in the 2D TOCSY spectrum.

It is interesting to note that the last two lysine residues are not structured. The mutation of Lys148 to Ala was reported to have no detectable effect on the interaction of DnaG-C with the DnaB helicase, while the substitution of Lys147 to Ala in full-length DnaG reduced DnaB priming activity about two-fold (**11**). The unordered structure of residues 1–10 is consistent with a comparison of sequences of primases related to that of the *E. coli* protein. Residues 429-433 at the C-terminus of the central domain were also disordered in crystals, suggesting that DnaG has a flexible linker of about 15 residues that connects the two domains. Flexibility can be required for the function in a primosome at the replication fork.

Although the present NOE analysis indicates that the long helical segment connecting the N- and C-terminal subdomains of DnaG-C (Fig. 23) is rigid on a sub-nanosecond timescale, there may be increased mobility on longer timescales. This could be analyzed by the measurement of amide proton exchange rates with the water. Possibly, however, the unexpected apparent rigidity of the connecting helix may also be due to homodimerization in solution which could be assessed by ultracentrifugation, gel filtration experiments or the observation of intermolecular NOEs.

REFERENCES

References

1. Y. Ohtomo, T. Bergman, B. L. Johansson, H. Jörnvall and J. Wahren, *Diabetologia*, 1998, 41, 287-291.
2. J. Wahren, K. Ekberg, B. Samnegård, and B. Johansson, *Curr. Diabet. Rep.*, 2001, 1, 261-266.
3. D. V. Waterhous and W. C. Johnson, *Biochemistry*, 1994, 33, 2121-2128.
4. L. R. Brown and K. Wuthrich, *Biochim. Biophys. Acta.*, 1981, 647, 95-111.
5. R. Schwyzer, *Biochemistry*, 1986, 25, 6335-6350.
6. M. R Tessmer and D. A. Kallick, *Biochemistry*, 1997, 36, 1971-1981.
7. M. Henriksson, J. Shafqat, E. Liepinsh, M. Tally, J. Wahren, H. Jörnvall and J. Johansson, *Cell. Molec. Life Sci.*, 2000, 57, 337-342.
8. Y. Wang and O. Jardetzky, *J. Am. Chem. Soc.*, 2002, 124, 14075-14084.
9. K. Wüthrich, *NMR of proteins and nucleic acids*, 1986, John Wiley & Sons. New York.
10. A. Kornberg and T. A. Baker, *DNA replication*, 1991, Freeman, New York.
11. K. Tougu and K. J. Marians, *J. Biol. Chem.*, 1996, 271, 21398-21405.
12. K. Loscha, A. J. Oakley, B. Bancia, P. Schaeffer, P. Prosselkov, G. Otting, M. C. J. Wilce and N. E Dixon, *Protein Expr. Purif.*, 2003, 33, 304-310.
13. N. J. P. Stamford, P. E. Lilley and N. E. Dixon, *Biochim. Biophys. Acta*, 1992, 1132, 17-25.
14. W. G. Proctor and F. C Yu, *Phys. Rev.*, 1950, 77, 717.
15. H. S. Gutowsky, D. W. McCall and C. P Schlicter, *Phys. Rev.*,1951, 84, 589-590.
16. A.W. Overhauser, *Phys. Rev.*, 1955, 92, 411-415.
17. R. R. Ernst and W. A. Anderson, *Rev. Sci. Inst.*, 1966, 37, 93-102.
18. W. P. Aue, E. Bartholdi and R. R. Ernst, *J. Chem. Phys.*, 1976, 64, 2229-2246.
19. K. Nagayama, K. Wuthrich, R. R. Ernst, *Biochem. Biophys. Res. Commun.*, 1977, 90, 305-311.
20. G. Wagner, K. Anil and K. Wuthrich, *Eur. J. Biochem.*, 1981, 114, 375-384.
21. G. Wagner and K. Wuthrich, *J. Mol. Biol.*, 1982, 155, 347-366.
22. K. Wuthrich, G. Wider, G. Wagner and W. Braun, *J. Mol. Biol.*, 1982, 155, 311-319.
23. M. P. Williamson, T. F. Havel and K. Wuthrich, *J. Mol. Biol.*, 1985, 182, 295-315.

24. L. E. Kay and K. H. Gardner, *Curr. Opin. Struct. Biol.,* 1997, 7, 722-731.

25. L. E. Kay, *Curr. Opin. Struct. Biol.*, 1995, 5, 674-681.

26. R. E. Hurd and B. K. John, *J. Magn. Reson.*, 1991, 92, 658-668.

27. G. W. Vuister, R. Boelens, R. Kaptein, M. Burgering and P. C. M. Van Zilj *J. Biomol. NMR*, 1992, 2, 301-305.

28. G. A. Morris and R. Freeman, *J. Am. Chem. Soc.*, 1979, 101, 760-762.

29. C. Griesinger, O. W. Sorensen and R. R. Ernst, *J. Am. Chem. Soc.*, 1985, 6394-6396.

30. L. Braunschweiler and R. R. Ernst, *J. Magn. Reson.*, 1983, 53, 521-528.

31. E. Kupce and R. Freeman, *J. Magn. Reson.*, 1995, 115, 273-276.

32. A. Bax and R. Freeman, *J. Am. Chem. Soc.*, 1981, 44, 542-561.

33. O. W. Sorensen, C. Griesinger and R. R. Ernst, *J. Am. Chem. Soc.*, 1985, 107, 7778-7779.

34. H. Geen and R. Freeman, *J. Magn. Reson.,* 1991, 93, 93 – 141.

35. M. Ikura, L. E. Kay and A. Bax, *Biochemistry*, 1990, 29, 4659-4667.

36. S. G. Hyberts and G. Wagner, *J. Biomol. NMR*, 2003, 26, 335-344.

37. G. Bodenhausen and D. J. Ruben, *Chem. Phys. Lett.*, 1980, 69, 185.

38. S. B. Shuker, P. J. Hajduc, R. P. Meadows and S. W. Fesik, *Science*, 1996, 274, 1531-1534.

39. L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting and N. Dixon, *FEBS Lett.*, 2002, 524, 159-162.

40. D. Marion, P. Driscoll, L. E. Kay, P. Wingfield, A. M. Gronenborn and G. M. Clore, *Biochemistry*, 1989, 28, 6150.

41. M. Wittekind and L. Muller, *J. Magn. Reson.*, 1992, 101B, 201-205.

42. E. Kupce and R. Freeman, *J. Magn. Reson.*, 1995, 115, 273-276.

43. L. E. Kay, M. Ikura, R. Tschudin and A. Bax, *J. Magn. Reson.*, 1990, 89, 496-514.

44. Stonehouse, R. T. Clowes, G. L. Shaw, J. Keller and E. D. Laue, *J. Biomol. NMR.,* 1995, 5, 226-232.

45. R. Ishimia and D. A. Torchia, *Nat. Struct. Biol.*, 2000, 7(9), 740-743.

46. N. J. Skelton, A. G. Palmer III, M. Akke, J. Kordel, M. Rance and W. J. Chazin, *J. Magn. Reson.*, 1992, 102(B), 253-264.

47. J. Boyd, *J. Magn. Reson.*, 1994, 107(B), 279-285.

48. N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E Shoelson, T. Powson, J. D. Forman-Kay and L. E. Kay, *Biochemistry*, 1994, 33, 5984-6003.

49. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 1982, 104, 4646-4559.

50. G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 1982, 104, 4559-4570.

51. A. Abragam, *Principles of Nuclear Magnetism*, 1961. Oxford University Press, London.

52. G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll and A. M. Gronenborn, *J. Am. Chem. Soc.*, 1990, 112, 4989-4991.

53. G. M. Clore, P. C. Driscoll, P. T. Wingfield and A. M. Gronenborn, *Biochemistry*, 1990, 29, 7387-7401.

54. A.W. Overhauser, *Phys. Rev.*, 1955, 92, 411-415.

55. V. Bushuev, A. T. Gudkov, A. Liljas and N. F. Sepetov, *J. Biol. Chem.*, 1989, 264, 4498-4505.

56. D. S. Wishart, B. D. Sykes and F. M. Richard, *Biochemistry*, 1992, 31, 1647-1651.

57. Y. Wang and O. Jardetzky, *J. Am. Chem. Soc.*, 2002, 124, 14075-14084.

58. D. Gust, R. B. Moon and J. D. Roberts, *Proc. Natl. Acad. Sci. USA.*, 1975, 72, 4696-4700.

59. M. Witanovski, L. Stefaniak and G.A Webb, *Annu. Rep. NMR Spect,* 1985, 18, 1.

60. B. Tyler, *Annu. Rev. Biochem.,* 1978, 47, 1127-1162.

61. F. C. Neidhardt, *Escherichia Coli and Salmonella typhimuria: Cellular Molecular Biology, Vol 1*, 1987, Amer. Soc. Microbiol., Washington D.C.

62. H.R. Kricheldorf, *Pure Appl. Chem.,* 1982, 54, 467.

63. A. M. Gronenborn and G. M. Clore, *Protein Sci.*, 1996, 5, 174-177.

64. Z. Serber and V. Dotsch, *Biochemistry*, 2001, 40, 14317-14323.

65. S. Kim, H. G. Dallamann, C. S McHenry and K. J. Marians, *Cell*, 1996, 84, 643-651.

66. H. D. Ou, H. C. Lai, Z. Serber and V. Dotsch, *J. Biomol. NMR*, 2001, 21, 269-273.

67. E. Kariya, S. Ohki, T. Hayano and M. Kainosho, *J. Biomol. NMR,* 2000, 18, 75-76.

68. P. M. Clark, *Ann. Clin. Biochem.*, 1999, 36, 541-564.

69. D. Neri, M. Billeter, G. Wider, and K. Wüthrich, *Science*, 1992, 257, 1559-1563.

70. F. Delaglio, S. Grzesiek,, G. Zhu, G. W. Vuister, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, 6, 277-293.

71. T. D. Goddrad and D. G. Kneller, *University of California*, Sparky 3.

72. R. Rigler, A. Pramanik, P. Jonasson, G. Kratz, O. Jansson, P. Nygren et al, *Proc. Natl. Acad. Sci. USA*, 1999, 96, 13318-13323.

73. H.E. Hamm, *J. Biol. Chem.*, 1998, 273, 669-672.

74. J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry - fifth edition*, 2001, W. H. Freeman & Co., New York.

75. G. C. Allen, N. E. Dixon and A. Kornberg, *Cell*, 1993, 74, 713-722.

76. C. M. Elvin, P. R. Thompson, M. E. Argall, P. Hendry, N. P. J. Stamford, P. E. Lilley and N. E. Dixon, *Gene*, 1990, 87, 123-126.

77. P. E. Lilley, N. P. J. Stamford, S. G. Vaudevan and N. E. Dixon, *Gene*, 1993, 129, 9-16.

78. C. A. Love, P. E. Lilley and N. E. Dixon, *Gene*, 1996, 176, 49-53.

79. S. Hamdan, S. E. Brown, P. R. Thompson, J. Y. Yang, P. D. Carr, D. L. Ollis, G. Otting and N. E. Dixon, *J. Struct. Biol.*, 2000, 131, 164-169.

80. G. C. Allen and A. Kornberg, *J. Biol. Chem.*, 1993, 268, 19204-19209.

81. E. L. Zechner, C. A. Wu and K. J. Marians, *J. Biol. Chem.*, 1992, 267, 4054-4063.

82. K.Tougu, H. Peng and K. J. Marians, *J. Biol. Chem.*, 1994, 269, 4675-4682.

83. D. Fass, C. E. Bogden and C. E. Berger, *Structure*, 1999, 7, 691-698.

84. J. Weigelt, S. E. Brown, C. S. Miles, N. E. Dixon and G. Otting, *Structure*, 1999, 7, 681-690.

85. I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani and A. Zoroddu, *Proc. Natl. Acad. Sci. USA,* 2004, 101, 6841-6846.

86. Y. S. Babu, C. E. Bugg and W. J. Cook, *J. Mol. Biol.,* 1988, 204, 191-204

87. R. T. Schumacher, A. F. Rivard, H. P. Bachinger and J. P. Adelman, *Nature,*2001, 410, 1120-1124

88. G. Barbato, M. Ykura, L. E. Kay, R. W. Pastor and A. Bax, *Biochemistry*, 31, 5269-5278

89. P. Güntert, V. Dötsch, G. Wider and K. Wüthrich, *J. Biomol. NMR,* 1992, 2, 19-29.

90. C. Bartels, T. H. Xia, M. Billeter, P. Güntert and K. Wüthrich, *J. Biomol. NMR,* 1995, 5, 1-10.

91. D. R. Muhandiram and L. E. Kay, *J. Magn. Reson.*, 1994, 103, 203-216.

92. H. Oschkinat, T. Muller and T. Dieckmann, *Angew. Chem. Int. Ed. Eng.*, 1994, 33, 277-293.

93. C. Renner, M. Schleicher, L. Moroder and T.A. Holak, 2002, *J. Biomol. NMR*, 23, 23-33.

94. L. E Kay, P. Keifer and T. Saarinen, *J. Am. Chem. Soc.*, 1992, 114, 10663-10665.

95. L. Kay, M. Wittekind, M. McCoy, M. Friedrichs, A. Bell, E. Ernst, T. Lavoie and L. Muller, *J. Magn. Reson.*, 1992, 98, 443-350.

APPENDIX I

NMR pulse sequences

Pulse sequences are shown for each of the NMR experiments performed with the exception of 2D DQF-COSY, 2D TOCSY and 2D NOESY which are very common experiments. Pulses labelled with pw (7.1 μs) represent a flip angle of 90^0, while pulses labelled with 2*pw (14.2 μs) correspond to a flip angle of 180^0. The Tx, Dec and Dec2 labels represent the ^1H, ^{13}C and ^{15}N spectrometer channels, respectively. Water suppression was achieved using one-lobe H_2O sinc pulses or WATERGATE and the ^1H carrier was placed on the water resonance for the duration of the experiment. Phase cycle and delays were used as described in literature. All gradient units are in mG/cm, while gradient delays are written on top of each gradient. Delaya are specified in milliseconds (m) or microseconds (u).

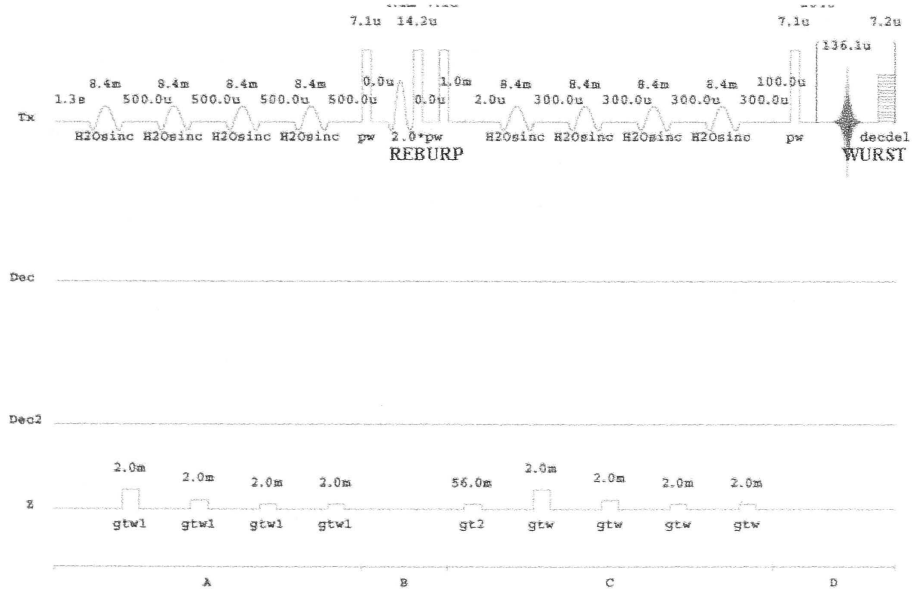


Fig. 1 2D NOESY with ω_1 and ω_2 decoupling. $^1H^{\alpha}$-selective decoupling was achieved using a WURST pulse during the acquisition and a REBURP pulse during the evolution period. The gradient strengths were: gtw1 (16000, 8000, 4000, 2000), gt2 (300) and gtw (16000, 8000, 4000, 2000) (**31**, **32**).

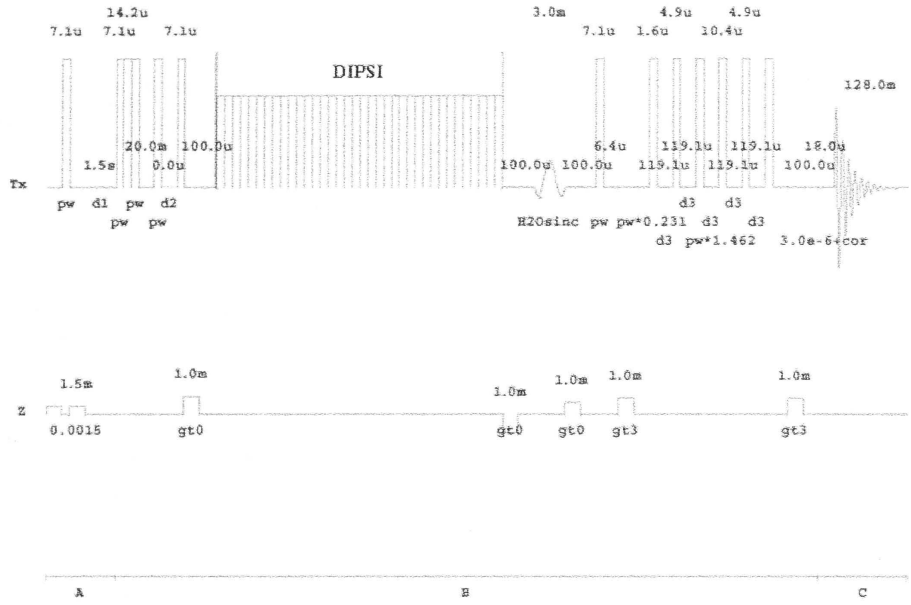


Fig. 2 2D TOCSY pulse sequence designed for T_1 relaxation measurements. A 90^0 composite pulse and a delay T were inserted at the beginning of a 2D TOCSY pulse sequence employing 3919-water suppression. The gradient strengths were: 1000 for the first two, gt0 (9000) and gt3 (8000) (**30**).

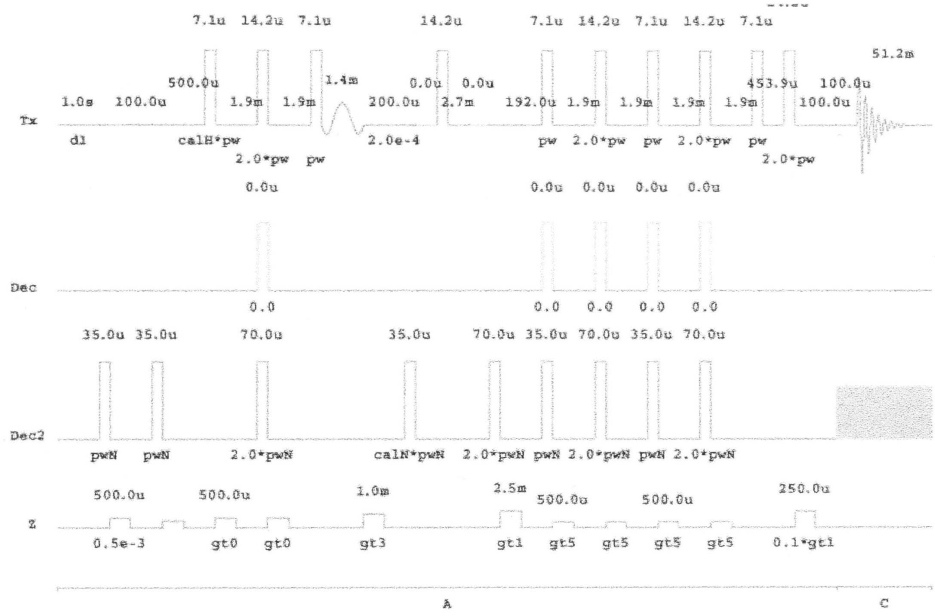


Fig. 3 2D ^{15}N-HSQC (sensitivity-enhanced). The gradient strengths were: gt0 (8000), gt3 (11000), gt1 (13000) and gt5 (2000) (**37**)

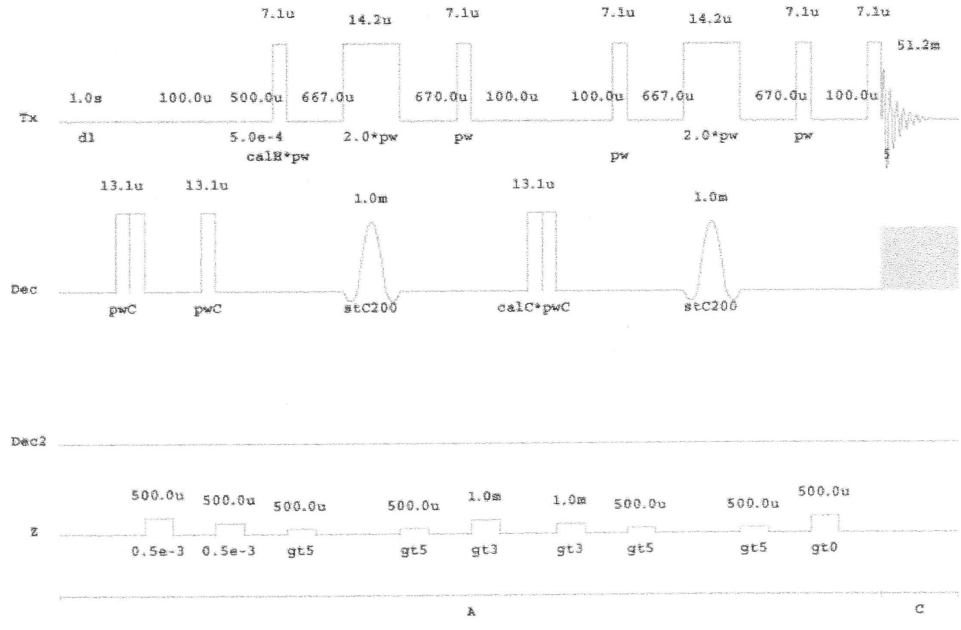


Fig. 4 2D ^{13}C-HSQC. The gradient strengths were: 13000 and 9100 for the first two, gt5 (2000), gt3 (11000) and gt0 (13000) (**94**)

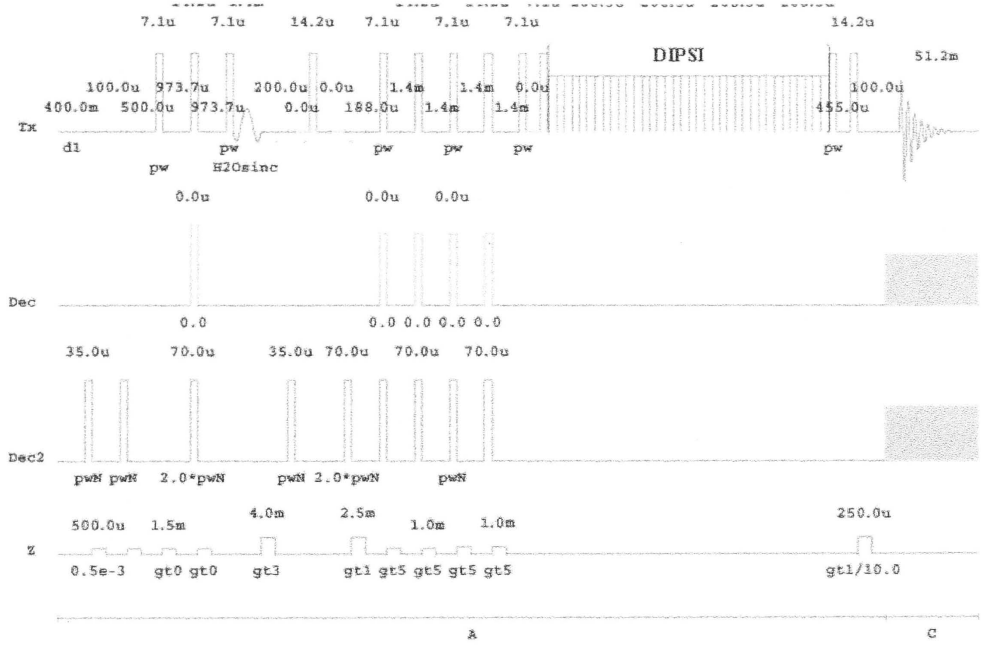


Fig. 5 2D TOCSY-relayed ^{15}N-HSQC (sensitivity-enhanced). The gradient strengths were: gt0(1400), gt3 (13000), gt1 (13000) and gt5 (3745) (**40**).

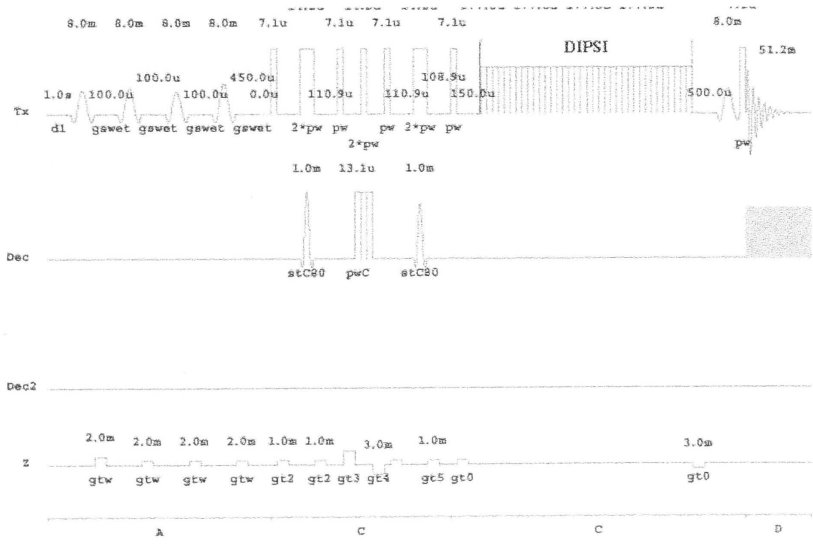


Fig. 6 2D TOCSY-relayed ^{13}C HSQC. The gradient strengths were: gtw (8000), gt2 (2106), gt3 (13000) and gt5 (3745) (**94**).

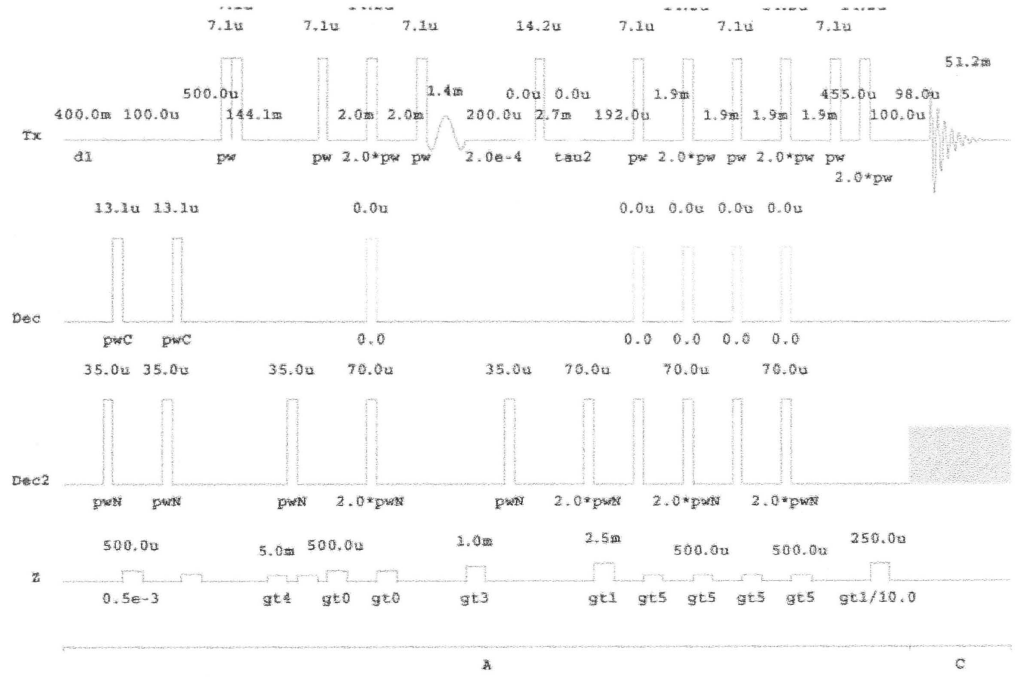


Fig. 7 Sensitivity-enhanced 3D NOESY-^{15}N -HSQC experiment. The gradient strengths were: gt4 (500), gt0 (8000), gt3 (11000), gt1 (13000) and gt5 (2000) (**40**)

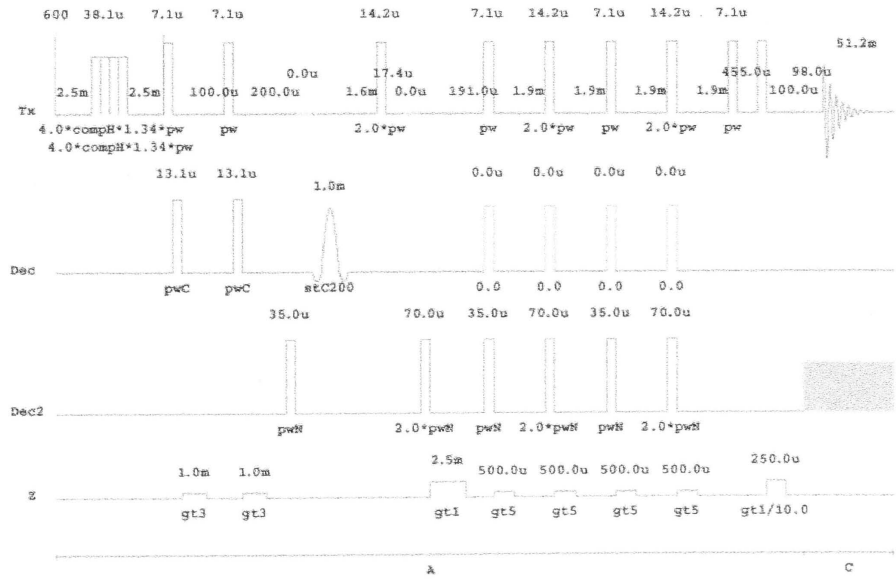


Fig. 8 2D {^{1}H}^{15}N NOE experiment. The gradient strengths were: gt3 (3000), gt1 (13000) and gt5 (2500) (**93**).

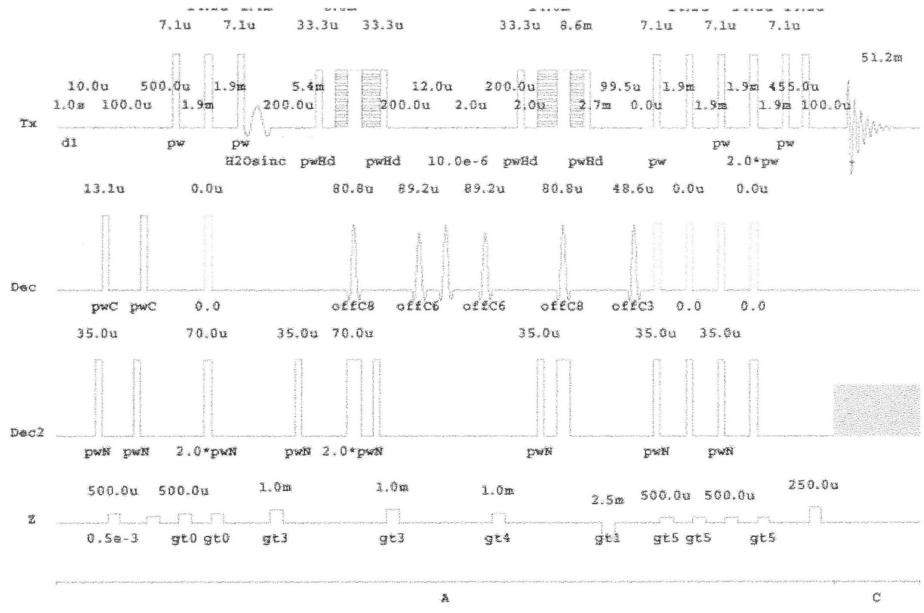


Fig. 9 Sensitivity-enhanced HNCO using pulsed field gradients to select for the coherence transfer pathway involving ^{15}N magnetization and employing a water selective pulse to minimize saturation of water. The gradient strengths were: gt0 (8000), gt3 (11000), gt4 (8000) gt5 (2000) and gt1 (13000) (**43**).

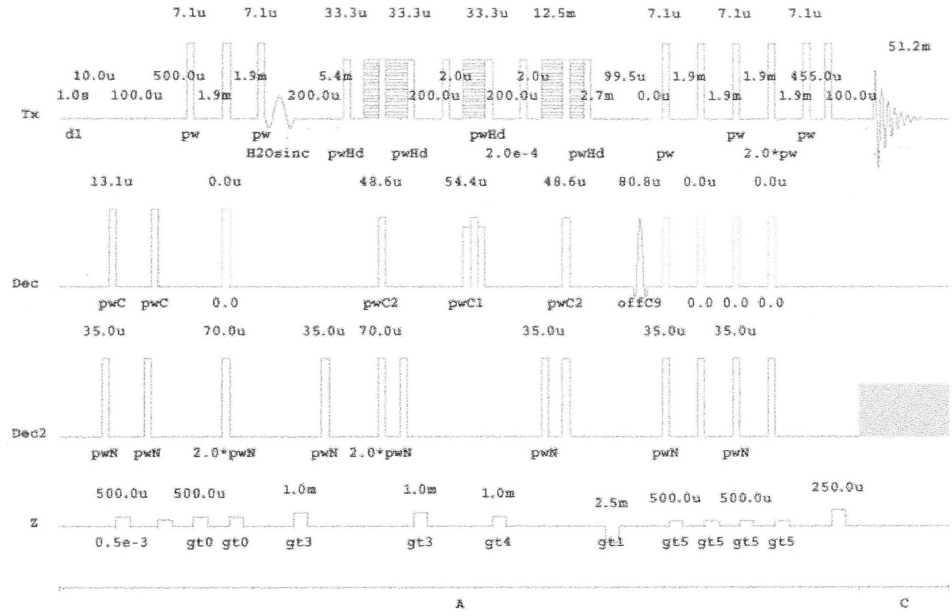


Fig. 10 Pulse sequence for the sensitivity-enhanced HNCA experiment employing a water selective pulse minimize saturation of water. The gradient strengths were: gt0 (8000), gt3 (11000), gt4 (8000), gt5 (2000) and gt1 (11000) (**35**).

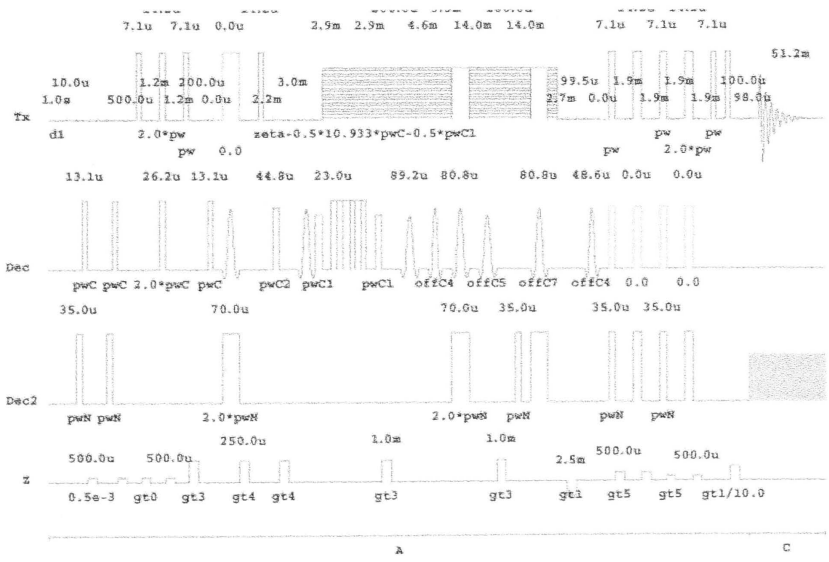


Fig. 11 Pulse sequence for the sensitivity-enhanced CBCA(CO)NH experiment. The gradient strengths were: gt0 (4000), gt3 (19000), gt4 (19000), gt5 (8000) and gt1 (13000) (**42**).

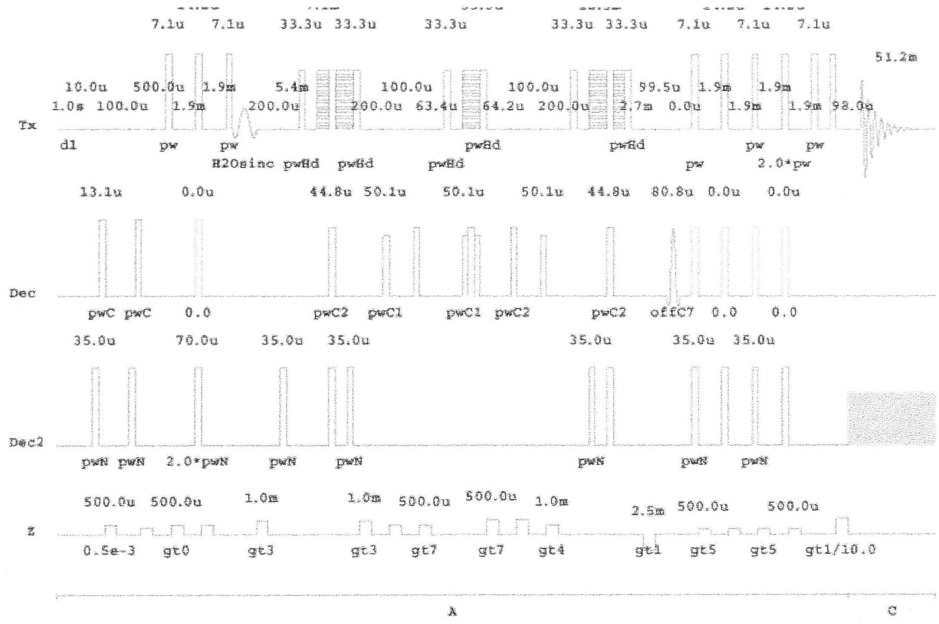


Fig. 12 Pulse sequence for the sensitivity-enhanced HNCACB experiment. The gradients strengths were: gt0 (8000), gt3 (11000), gt7 (12000), gt4 (12000), gt5 (2000) and gt1 (13000) (**91**).

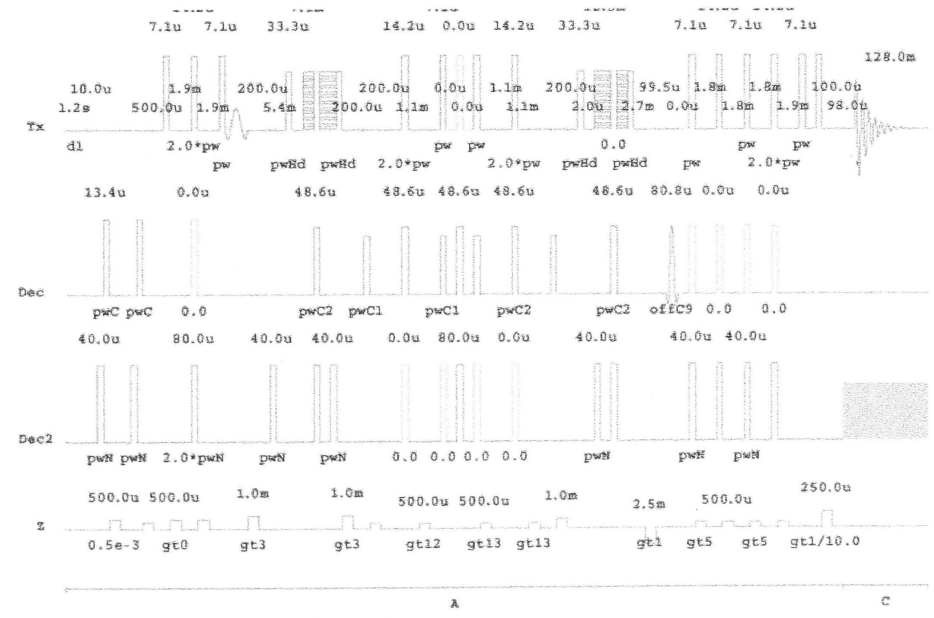


Fig. 13 Pulse sequence for the sensitivity-enhanced HNCAHA experiment. The gradient strengths were: gt0 (8000), gt3 (11000), gt12 (1000), gt13 (7500), gt5 (2000) and gt1 (13000) (**95**).

APPENDIX II

C-peptide

C-program for chemical shift prediction in random coil peptides

```
/* Program for predicting NMR amino acid chemical shifts.
 * Array x stores a sequence of amino acid character codes read from a file.
 * This program sums the averaged random coil chemical shift from x[i]
 * plus the averaged neighboring residue effects from x[i-1] and x[i+1]
 * to give the predicted chemical shift of x[i].
 */

#include<stdio.h>
#include<ctype.h>
#define SIZE 200

/*
predicted chemical shift:

sum=x[i-1] + x[i] + x[i+1]

HA: haa, hba, hca
15N: na, nb, nc
CA: caa, cab, cac
CB:  cba, cbb, cbc
NH: nha, nhb, nhc
CO: coa, cob, coc

*/

main()

{
 char x[SIZE], d, c;
 int i, N;
 double haa, hba, hca, na, nb, nc, caa, cab, cac, cba, cbb, cbc, nha, nhb, nhc, coa, cob, coc;
 double sum;

 i=0;
 while( (c=getchar()) != EOF){
  if(isspace(c))
    continue;
  /*printf("Char = %c\n",d);*/
  x[i]=c;
  /*    printf("i=%d, x[i]=%c\n", i, x[i]);*/
  ++i;
  if (i >= SIZE)
    printf("error: resize array\n");
 }


 N=i;

 for(i=0;  i < N; ++i){
  sum=0;
 }


 for(i=1; i<N-1; ++i){
  switch(x[i]){
  case 'A':
    hba=4.31;
```

```
    na=123.82;
    caa=52.46;
    cba=18.98;
    nha=8.09;
    coa=177.28;
    break;
case 'D':
    hba=4.62;
    na=120.37;
    caa=54.00;
    cba=40.78;
    nha=8.31;
    coa=176.00;
    break;
  case 'R':
    hba=4.26;
    na=120.75;
    caa=56.18;
    cba=30.36;
    nha=8.21;
    coa=176.01;
    break;
  case 'C':
    hba=4.59;
    na=118.10;
    caa=58.24;
    cba=29.54;
    nha=8.10;
    coa=175.11;
    break;
  case 'N':
    hba=4.66;
    na=118.50;
    caa=53.00;
    cba=38.43;
    nha=8.35;
    coa=174.84;
    break;
  case 'Q':
    hba=4.29;
    na=119.82;
    caa=55.89;
    cba=29.01;
    nha=8.20;
    coa=175.75;
    break;
  case 'E':
    hba=4.28;
    na=120.62;
    caa=56.66;
    cba=29.87;
    nha=8.36;
    coa=176.32;
    break;
  case 'G':
    hba=3.97;
    na=109.48;
    caa=45.28;
    cba=0;
    nha=8.37;
```

```
    coa=174.01;
    break;
  case 'H':
    hba=4.60;
    na=118.92;
    caa=55.74;
    cba=29.50;
    nha=8.18;
    coa=174.78;
    break;
  case 'I':
    hba=4.18;
    na=120.58;
    caa=60.79;
    cba=38.43;
    nha=7.94;
    coa=175.52;
    break;
  case 'L':
    hba=4.36;
    na=121.57;
    caa=54.77;
    cba=42.14;
    nha=8.06;
    coa=176.70;
    break;
  case 'K':
    hba=4.28;
    na=121.10;
    caa=56.29;
    cba=32.53;
    nha=8.17;
    coa=176.15;
    break;
  case 'M':
    hba=4.47;
    na=120.14;
    caa=55.43;
    cba=32.92;
    nha=8.22;
    coa=175.94;
    break;
  case 'F':
    hba=4.59;
    na=119.72;
    caa=57.46;
    cba=39.41;
    nha=8.09;
    coa=175.46;
    break;
  case 'P':
    hba=4.41;
    na=0;
    caa=63.24;
    cba=31.81;
    nha=0;
    coa=176.62;
    break;
  case 'S':
    hba=4.45;
```

```
  na=116.00;
  caa=58.20;
  cba=63.75;
  nha=8.22;
  coa=174.41;
  break;
 case 'T':
  hba=4.44;
  na=113.88;
  caa=61.30;
  cba=68.92;
  nha=8.16;
  coa=174.78;
  break;
 case 'W':
  hba=4.60;
  na=120.99;
  caa=57.54;
  cba=29.60;
  nha=7.97;
  coa=175.87;
  break;
 case 'Y':
  hba=4.56;
  na=119.37;
  caa=57.64;
  cba=38.78;
  nha=7.99;
  coa=175.29;
  break;
 case 'V':
  hba=4.13;
  na=119.91;
  caa=62.00;
  cba=32.35;
  nha=7.98;
  coa=175.66;
  break;
 default:
 printf("\n\nWarning: %c is not a valid code.\n\n", x[i]);
 break;
 }


 switch(x[i-1]){
 case 'A':
  haa=-0.01;
  nb=-2.21;
  cab=-0.01;
  cbb=-0.04;
  nhb=-0.07;
  cob=0.14;
  break;
 case 'D':
  haa=-0.03;
  nb=-0.01;
  cab=0.20;
  cbb=-0.07;
  nhb=-0.01;
  cob=0.07;
```

```
    break;
  case 'R':
    haa=0.01;
    nb=-0.45;
    cab=0.00;
    cbb=0.01;
    nhb=-0.03;
    cob=-0.07;
    break;
  case 'C':
    haa=0;
    nb=1.36;
    cab=0.44;
    cbb=-0.18;
    nhb=0.03;
    cob=0.00;
    break;
  case 'N':
    haa=-0.03;
    nb=-0.76;
    cab=0.18;
    cbb=-0.20;
    nhb=0.01;
    cob=-0.19;
    break;
  case 'Q':
    haa=0;
    nb=-0.09;
    cab=0.13;
    cbb=0.06;
    nhb=0.01;
    cob=0.07;
    break;
  case 'E':
    haa=-0.01;
    nb=-0.36;
    cab=0.01;
    cbb=0.04;
    nhb=-0.01;
    cob=-0.14;
    break;
  case 'G':
    haa=0.04;
    nb=-0.43;
    cab=-0.26;
    cbb=0.17;
    nhb=-0.10;
    cob=-0.09;
    break;
  case 'H':
    haa=-0.05;
    nb=-0.05;
    cab=0.16;
    cbb=0.07;
    nhb=-0.01;
    cob=-0.13;
    break;
  case 'I':
    haa=0.04;
    nb=2.92;
```

```
    cab=-0.15;
    cbb=0.26;
    nhb=0.12;
    cob=0.11;
    break;
  case 'L':
    haa=0;
    nb=-0.76;
    cab=-0.07;
    cbb=-0.03;
    nhb=-0.06;
    cob=0.13;
    break;
  case 'K':
    haa=0;
    nb=-0.26;
    cab=-0.07;
    cbb=0.08;
    nhb=-0.02;
    cob=-0.09;
    break;
  case 'M':
    haa=0.05;
    nb=0.69;
    cab=0.09;
    cbb=0.10;
    nhb=0.05;
    cob=0.10;
    break;
  case 'F':
    haa=-0.01;
    nb=0.20;
    cab=0.02;
    cbb=-0.06;
    nhb=0.01;
    cob=-0.14;
    break;
  case 'P':
    haa=-0.04;
    nb=-0.94;
    cab=0.07;
    cbb=-0.16;
    nhb=0.14;
    cob=0.21;
    break;
  case 'S':
    haa=0.01;
    nb=1.16;
    cab=0.11;
    cbb=-0.06;
    nhb=0.02;
    cob=-0.10;
    break;
  case 'T':
    haa=0;
    nb=1.23;
    cab=0.05;
    cbb=-0.11;
    nhb=0.02;
    cob=-0.07;
```

```
  break;
 case 'W':
  haa=-0.05;
  nb=0.97;
  cab=0.00;
  cbb=0.28;
  nhb=-0.08;
  cob=-0.46;
  break;
 case 'Y':
  haa=-0.01;
  nb=0.46;
  cab=-0.19;
  cbb=-0.10;
  nhb=0.02;
  cob=-0.03;
  break;
 case 'V':
  haa=0.04;
  nb=2.77;
  cab=-0.16;
  cbb=0.03;
  nhb=0.17;
  cob=0.00;
  break;
 default:
 if (i==1)
 printf("\n\nWarning: %c is not a valid code.\n\n", x[0]);
 break;
 }


 switch(x[i+1]){
 case 'A':
  hca=-0.03;
  nc=-0.11;
  cac=0.07;
  cbc=-0.09;
  nhc=-0.01;
  coc=0.05;
  break;
 case 'D':
  hca=-0.03;
  nc=0.23;
  cac=0.28;
  cbc=0.11;
  nhc=0.04;
  coc=-0.11;
  break;

 case 'R':
  hca=-0.02;
  nc=-0.09;
  cac=0.19;
  cbc=-0.12;
  nhc=-0.01;
  coc=0.16;
  break;
 case 'C':
  hca=0.03;
```

```
    nc=-1.17;
    cac=0.17;
    cbc=0.21;
    nhc=0.01;
    coc=0.10;
    break;
  case 'N':
    hca=-0.04;
    nc=-0.03;
    cac=0.24;
    cbc=-0.06;
    nhc=0.04;
    coc=-0.23;
    break;
  case 'Q':
    hca=-0.05;
    nc=-0.31;
    cac=0.37;
    cbc=-0.13;
    nhc=0.01;
    coc=0.10;
    break;
  case 'E':
    hca=-0.04;
    nc=-0.26;
    cac=0.25;
    cbc=0.06;
    nhc=0.07;
    coc=0.14;
    break;
  case 'G':
    hca=-0.03;
    nc=0.13;
    cac=0.12;
    cbc=-0.07;
    nhc=0.06;
    coc=0.47;
    break;
  case 'H':
    hca=-0.09;
    nc=-0.09;
    cac=0.22;
    cbc=-0.24;
    nhc=-0.01;
    coc=-0.05;
    break;
  case 'I':
    hca=0.03;
    nc=-0.20;
    cac=0.03;
    cbc=0.28;
    nhc=-0.01;
    coc=-0.09;
    break;
  case 'L':
    hca=-0.02;
    nc=-0.49;
    cac=0.10;
    cbc=-0.10;
    nhc=-0.07;
```

```
    coc=0.06;
    break;
  case 'K':
    hca=-0.03;
    nc=-0.13;
    cac=0.08;
    cbc=0.01;
    nhc=-0.04;
    coc=-0.13;
    break;
  case 'M':
    hca=-0.01;
    nc=-0.02;
    cac=0.22;
    cbc=0.06;
    nhc=0.01;
    coc=0.19;
    break;
  case 'F':
    hca=-0.02;
    nc=-0.35;
    cac=-0.04;
    cbc=-0.09;
    nhc=-0.04;
    coc=-0.22;
    break;
  case 'P':
    hca=0.21;
    nc=0.92;
    cac=-2.04;
    cbc=-0.20;
    nhc=-0.17;
    coc=-1.19;
    break;
  case 'S':
    hca=0.02;
    nc=0.30;
    cac=0.10;
    cbc=0.14;
    nhc=0.07;
    coc=0.10;
    break;
  case 'T':
    hca=0.08;
    nc=0.22;
    cac=0.02;
    cbc=0.14;
    nhc=0.04;
    coc=0.10;
    break;
  case 'W':
    hca=-0.03;
    nc=-0.59;
    cac=-0.06;
    cbc=-0.03;
    nhc=-0.10;
    coc=-0.33;
    break;
  case 'Y':
    hca=-0.06;
```

```
    nc=-0.48;
    cac=0.10;
    cbc=0.00;
    nhc=-0.07;
    coc=-0.51;
    break;
  case 'V':
    hca=0.05;
    nc=-0.05;
    cac=0.05;
    cbc=0.18;
    nhc=0.01;
    coc=-0.08;
    break;
  default:
    if (i== N-2)
    printf("\n\nWarning: %c is not a valid code.\n\n", x[N-1]);
    break;

  }

 sum = (haa + hba + hca);


  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i,haa,hba,hca);*/
  printf("  HA  %.2f\n", sum);


/* Glycine doesn't have CB */
  sum=(cba + cbb + cbc );
  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i, na, nb, nc);*/
  if ((sum > -1) && (sum < 1))
  printf("  CB   -\n");
  else
  printf("  CB  %.2f\n", sum);



sum=(caa + cab +cac );
  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i, caa, cab, cac);*/

  printf("  CA  %.2f\n", sum);



 /* Proline doesn't have HN  and 15N  */


sum=(na + nb +nc );

  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i, na, nb, nc);*/
if ((sum > -2.78) && (sum < 2.78))
  printf("  15N   -\n");
  else
```

```
  printf(" 15N  %.2f\n", sum);


  sum=(nha + nhb +nhc );
  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i, na, nb, nc);*/
if ((sum > -1) && (sum < 1))
  printf("  NH  -\n");
  else

  printf("  HN  %.2f\n", sum);


 sum=(coa + cob +coc );
  printf("%c%2d", x[i], i+1);
  /*printf("%d, %f, %f, %f", i, na, nb, nc);*/
  printf("  CO  %.2f\n", sum);


 }


 }
```

Measured T_1 relaxation times (in seconds) of C-peptide in H_2O at pH 6 and 25 ^0C

Residue/Atom	1HA	HB	HB1	HB2	HD1	HD2	HG	HG1	HG2	HN
Glu1	0.800		0.437	0.520				0.624	0.624	-
Ala2	1.070	0.454								0.734
Glu3	0.791		0.409	0.409				0.511	0.511	0.526
Asp4	1.150		0.657	0.894						0.681
Leu5	0.973									0.668
Gln6	0.847		0.451	0.432				0.526	0.526	0.433
Val7	1.020						0.494			0.617
Gly8	0.625									0.577
Gln9	1.070		0.409	1.06				0.445	0.379	0.620
Val10	0.998						0.445			0.629
Glu11	0.943		0.467	0.472				0.501	0.501	0.525
Leu12	0.956									0.559
Gly13	0.618									0.557

Gly14	0.654								0.589
Gly15	0.612								0.647
Pro16			0.344	0.441	0.262	0.262	0.410		
Gly17	0.601								0.512
Ala18	1.091	0.475							0.717
Gly19	0.623								0.557
Ser20	1.104	0.555							0.541
Leu21	0.974								0.620
Gln22	0.897		0.504	0.383			0.498	0.498	0.705
Pro23		0.368	0.395	0.348	0.367	0.432			
Leu24	0.922								0.633
Ala25	1.070	0.454							0.734
Leu26									
Glu27	0.800		0.371	0.396			0.590	5.590	0.619
Gly28	0.623								0.557
Ser29	1.029	0.500							0.543
Leu30	0.784								0.476
Gln31	0.957		0.488	0.477			0.502	0.502	0.623

APPENDIX III

DnaG-C

Assignment for the $^{1}H^{N}$ and ^{15}N resonances of DnaG-C] in H_2O at pH 6 and 36 ^{0}C

Amino acid	w1 (^{15}N) (ppm)	w2 ($^{1}H^{N}$) (ppm)
E3	120.611	8.505
S4	116.665	8.315
G5	110.786	8.383
V6	118.618	7.856
S7	119.521	8.297
R8	123.899	8.199
V10	121.556	8.159
Q12	120.800	8.362
L13	123.199	8.140
R15	122.724	8.468
T16	114.421	7.660
T17	118.100	8.273
R19	118.366	7.571
L21	119.257	7.766
I22	116.495	8.237
G23	105.440	8.619
L24	119.987	8.805
L25	120.442	7.963
V26	114.422	8.248
Q27	113.884	7.533
N28	115.593	7.574
E30	117.181	9.173
L31	120.457	8.352
A32	119.274	7.613
T33	105.648	7.249
L34	120.202	7.841
V35	120.026	7.271
L38	122.724	8.468
E39	120.629	8.255
N40	115.097	8.600
L41	119.273	7.260
D42	121.347	8.247
E43	125.075	8.865
N44	116.456	8.491
K45	116.726	7.570
L46	117.194	6.822
G48	113.091	8.570
L49	122.958	8.426
G50	106.744	9.057
L51	124.357	7.786
F52	119.032	8.393
R53	116.973	8.845
E54	119.745	7.788
V56	121.102	8.520
N57	121.102	8.520

T58	121.102	8.520
C59	119.756	8.250
L60	117.436	8.294
S61	112.140	7.545
Q62	121.127	7.205
G64	111.345	8.696
L65	122.315	7.909
T66	110.705	7.770
T67	118.614	8.876
G68	106.115	8.641
Q69	120.916	7.693
L70	123.485	8.570
L71	117.202	8.490
H73	115.806	7.682
Y74	113.884	7.533
R75	124.827	7.518
G76	111.653	8.997
T77	108.857	7.716
N78	118.612	9.194
N79	114.307	7.882
A80	123.632	7.492
A81	118.996	8.345
T82	115.357	7.698
L83	119.745	7.788
E84	119.050	8.469
L86	118.185	7.975
S87	114.403	8.182
M88	118.826	7.157
W89	122.064	7.271
D90	126.285	8.529
D91	117.894	7.439
I92	112.810	6.843
A93	124.444	7.467
D94	118.124	7.686
K95	126.618	8.664
N96	116.721	8.573
I97	113.294	7.370
A98	127.307	7.664
E99	117.444	8.573
Q100	119.756	8.250
T101	115.799	8.263
F102	125.280	8.925
A103	113.082	8.678
S105	116.485	8.112
L106	124.448	8.403
N107	117.153	7.836
H108	117.458	7.885
Q110	117.245	8.184
D111	120.457	8.352
S112	115.822	8.202
T113	122.026	7.778

D114	121.019	7.878
R117	120.916	7.693
Q118	120.670	8.443
E119	118.094	8.198
E120	120.679	8.069
L121	120.652	8.206
I122	120.800	8.362
A123	122.350	7.935
R124	119.283	8.114
T127	109.814	7.809
H128	120.457	8.352
G129	109.312	8.128
L130	118.843	8.235
E133	119.138	8.794
E134	121.311	7.854
R135	120.172	8.470
L136	120.456	7.744
E137	122.957	8.039
L138	120.611	8.505
W139	119.521	8.297
T140	116.210	8.349
L141	123.620	8.453
N142	117.444	8.573
Q143	118.124	7.686
E144	119.723	8.064
L145	118.843	8.235
A146	121.817	7.467
K147	119.048	7.654
K148	126.922	7.620

^{15}N NOE values of DnaG-C in H_2O at pH 6 and 36 ^0C

Residue number	^{15}N-NOE (sec)	S/N (%)
3	-1.63	25
4	-4.04	21
5	-7.01	14
6	-1.36	25
7		
8	-1.13	15
9		
10	0.74	18
11		
12	0.72	32
13	1.00	13
14		
15	0.75	17
16		
17		
18		

19	0.72	18
20	0.65	25
21		
22	0.82	18
23	0.77	14
24	0.58	8
25	0.47	9
26	0.63	15
27	0.76	26
28		
29		
30	0.80	17
31	0.72	32
32	1.00	29
33	0.66	17
34	0.86	22
35	0.91	12
36		
37		
38	0.75	17
39	0.51	16
40	0.73	16
41	0.70	20
42		
43	0.73	13
44	0.67	24
45	0.73	21
46	0.80	16
47		
48	0.46	6
49	0.86	16
50	0.89	20
51	0.88	16
52	0.81	20
53	0.78	25
54	0.74	32
55	0.95	29
56	0.95	29
57	0.77	16
58	0.95	29
59	0.72	32
60	0.83	23
61	0.71	19
62	0.73	19
63		
64	0.61	16
65	0.71	16
66	0.78	15
67	1.05	15
68	0.87	15

69	1.15	18
70	1.03	16
71	1.03	10
72	0.65	25
73		
74	0.76	26
75	0.74	16
76	1.12	8
77	0.75	6
78		
79	0.62	17
80	0.81	8
81	0.81	21
82	0.91	25
83	0.74	32
84	0.68	24
85	0.79	15
86	0.95	16
87		
88	0.74	16
89	0.59	13
90	0.84	15
91	0.74	16
92	0.91	11
93	0.62	14
94	0.55	22
95	0.60	7
96	0.72	22
97	0.75	12
98	0.80	16
99	0.69	28
100	0.72	28
101		
102	0.79	12
103	0.92	14
104	0.84	20
105	0.76	19
106	0.83	15
107		
108	0.63	19
109		
110	0.84	15
111		
112	0.83	18
113	1.24	14
114	0.70	19
115	0.68	24
116	0.63	27
117		
118	0.62	21

119		
120		
121	0.84	14
122	0.72	32
123	0.71	15
124		
125	0.82	9
126		
127	0.58	8
128		
129		
130	0.57	26
131		
132		
133	0.70	11
134	0.70	11
135	0.68	25
136	0.82	11
137		
138		
139		
140	0.76	21
141		
142		
143	0.55	22
144	0.63	27
145		
146	0.39	9
147	0.20	9
148	-0.26	9

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 33 (2004) 304–310
Protein Expression & Purification
www.elsevier.com/locate/yprep

Expression, purification, crystallization, and NMR studies of the helicase interaction domain of *Escherichia coli* DnaG primase

Karin Loscha,[a] Aaron J. Oakley,[b] Bogdan Bancia,[a] Patrick M. Schaeffer,[a] Pavel Prosselkov,[a] Gottfried Otting,[a] Matthew C.J. Wilce,[b] and Nicholas E. Dixon[a,*]

[a] *Research School of Chemistry, Australian National University, Canberra ACT 0200, Australia*
[b] *School of Medicine and Pharmacology, and School of Biomedical and Chemical Sciences, University of Western Australia, Nedlands, WA 6907, Australia*

Received 8 September 2003, and in revised form 3 October 2003

Abstract

In *Escherichia coli*, the DnaG primase is the RNA polymerase that synthesizes RNA primers at replication forks. It is composed of three domains, a small N-terminal zinc-binding domain, a larger central domain responsible for RNA synthesis, and a C-terminal domain comprising residues 434–581 [DnaG(434–581)] that interact with the hexameric DnaB helicase. Presumably because of this interaction, it had not been possible previously to express the C-terminal domain in a stably transformed *E. coli* strain. This problem was overcome by expression of DnaG(434–581) under control of tandem bacteriophage λ-promoters, and the protein was purified in yields of 4–6 mg/L of culture and studied by NMR. A TOCSY spectrum of a 2 mM solution of the protein at pH 7.0, indicated that its structured core comprises residues 444–579. This was consistent with sequence conservation among most-closely related primases. Linewidths in a NOESY spectrum of a 0.5 mM sample in 10 mM phosphate, pH 6.05, 0.1 M NaCl, recorded at 36 °C, indicated the protein to be monomeric. Crystals of selenomethionine-substituted DnaG(434–581) obtained by the hanging-drop vapor-diffusion method were body-centered tetragonal, space group $I4_122$, with unit cell parameters $a = b = 142.2$ Å, $c = 192.1$ Å, and diffracted beyond 2.7 Å resolution with synchrotron radiation.

Keywords: DNA replication; DnaG primase; DnaB helicase; Primase–helicase interaction; Protein structure

The DnaG primase is a single-stranded DNA-dependent RNA polymerase and is an essential component of the replisome, the machine that accomplishes chromosomal DNA replication in *Escherichia coli* [1]. It is the enzyme responsible for repeated synthesis of short oligoribonucleotide primers that are extended by DNA polymerase III holoenzyme to produce Okazaki fragments on the lagging strand at replication forks, and it also synthesizes the initiating RNA primers on both leading strands at the origin of replication (*oriC*). It associates with other components of priming complexes, or primosomes [2], via interaction with the hexameric replicative helicase DnaB [3,4].

Sequences of eubacterial primases in the DnaG family (~65 kDa) are conserved [5] and have three separable domains [4,6–8]. Although there is no structure available of an intact primase from this family, crystal structures of the N-terminal 12 kDa domain of the *Bacillus stearothermophilus* primase [9] and the central catalytic core domain of *E. coli* DnaG [7,8] are known. The N-terminal domain is involved in template DNA recognition [10] and contains a zinc-binding motif [11] of the zinc ribbon subfamily of structures [9]. The central 36 kDa domain (residues 111–433) may be subdivided into three smaller sub-domains [6–8]. The first two of these are structurally related to the corresponding *TOPRIM* sub-domains in the structure of the primase moiety of the covalently linked bacteriophage T7 primase–helicase [12], and are together responsible for RNA synthesis [7,8,12].

The structure of the small C-terminal domain of the *E. coli* enzyme, here designated DnaG(434–581), is not known. DnaG(434–581) has previously been produced

* Corresponding author. Fax: +61261250750.
E-mail address: dixon@rsc.anu.edu.au (N.E. Dixon).

1046-5928/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2003.10.001

in vivo following phage infection of cells containing a plasmid with gene expression under control of a T7-promoter, but attempts to express the intact domain in a stably transformed *E. coli* host strain were unsuccessful [6]. This is presumed to be because DnaG(434–581) interferes profoundly with host DNA replication [13]. Effects of purified DnaG(434–581) on DnaB-dependent, but not DnaB-independent, in vitro DNA replication reactions showed that DnaG(434–581) bears the site of interaction of primase with the helicase [6], and a series of mutants were used to show that certain residues among the C-terminal 16 amino acids of DnaG are involved in interaction with DnaB [3,13]. The site of DnaG interaction in *E. coli* DnaB has not been precisely defined, but appears to involve the N-terminal domain of the helicase [14], whose structure has been determined independently by NMR [15] and crystallographic [16] techniques. Studies with the *B. stearothermophilus* enzymes suggest that there are sites for primase interaction in both the N- and C-terminal domains of the helicase [4].

As a further step towards understanding the structural basis for helicase–primase interaction in the bacterial replisome, we report here the construction of a stable host strain for production of DnaG(434–581), as well as purification and crystallization of the protein. Preliminary crystallographic and NMR data are presented.

Materials and methods

Construction of plasmids

Escherichia coli strain AN1459 [17] was used routinely as host strain during construction of plasmids. Plasmid pPL195 [11], which contains the complete *E. coli dnaG* gene (and some downstream chromosomal DNA), was used as template for a polymerase chain reaction, with the 5′-primer (5′-GAAGGAGATATA CAT*ATG*GCGGCAGAGAGCGGCGTTTC) designed to incorporate a start codon (italicized) before codon 434 of *dnaG* as part of an *Nde*I site (underlined). The other primer was the complement of a vector sequence beyond the *Eco*RI site downstream of *dnaG* in pPL195 [18]. The region encoding DnaG(434–581) contains a further *Nde*I site, so the amplified fragment was digested completely with *Eco*RI and partially with *Nde*I. The 694-bp fragment, isolated by agarose gel electrophoresis, was inserted between the same two sites in the bacteriophage λ-promoter vector pND706 [19] to yield pKL1176 (Fig. 1), and in the phage T7-promoter vector pETMCSI [20] to yield pKL1177 (5337 bp, not shown). The integrity of the inserted DNA in both plasmids was confirmed by determination of its nucleotide sequence using vector primers [18–20].

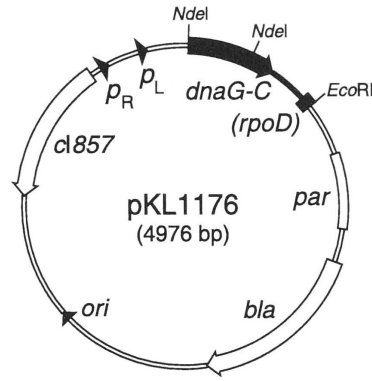


Fig. 1. Plasmid pKL1176, which directs overproduction of the C-terminal domain of *E. coli* DnaG primase, is a plasmid constructed by insertion of a PCR-generated fragment of *dnaG* (labeled *dnaG-C*) downstream of the tandem bacteriophage λp_R and p_L promoters in vector pND706 [19]. The fragment retains the 5′ end of the *rpoD* gene.

Overproduction of DnaG(434–581)

Strain BL21(DE3)*recA*/pKL1176 was grown in Terrific Broth [21] containing ampicillin at 50 mg/L at 30 °C in 1 L cultures until $A_{595} = 0.5$ and then shaken for a further 2 h at 42 °C. The cells were harvested by centrifugation and stored frozen at −70 °C. For preparation of selenomethionine-substituted (SeMet-) DnaG(434–581), cells were grown in the medium described by van Duyne et al. [22], and uniformly labeled [^{13}C,^{15}N]DnaG(434–581) was similarly prepared from cells grown in *E. coli* OD 2 CN medium (Silantes GmbH, Munich, Germany).

Purification of DnaG(434–581)

Cells from 5 L of culture (18 g) were thawed, resuspended in 270 ml lysis buffer (50 mM Tris–HCl, pH 7.6, 10% [w/v] sucrose, 150 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, and 10 mM spermidine · 3HCl), and lysed by being passed twice through a French press operating at 12,000 psi. All following steps were carried out at 0–4 °C. The lysate was centrifuged for 1 h at 38,000*g*. Proteins were precipitated from the supernatant (Fraction I, 280 ml, 1546 mg of protein) by addition of ammonium sulfate (0.30 g/ml) and collected by centrifugation for 45 min at 38,000*g*. The pellet was dissolved in 50 ml buffer A (50 mM Tris–HCl, pH 7.6, 1 mM dithiothreitol, 1 mM EDTA, and 10% v/v glycerol) containing 150 mM NaCl, and dialyzed against two changes of 1 L buffer A + 20 mM NaCl to yield Fraction II (692 mg protein). After centrifugation and dilution to 100 ml, the supernatant was applied at 1 ml/min to a column (14.5 × 2.5 cm) of Toyopearl TSK DEAE-650M anion-exchange resin. DnaG(434–581) eluted in a sharp peak at ~70 mM NaCl in a gradient (360 ml) of 20–250 mM NaCl in buffer A. Fractions containing DnaG(434–581) (Fraction III, 52 ml, 118 mg

protein) were pooled, dialyzed against two changes of 1 L buffer B (50 mM Tris–HCl, pH 7.6, 1 mM dithiothreitol, and 1 mM EDTA) containing 20 mM NaCl, and clarified by centrifugation. Portions (10 ml each) were loaded separately on a column of MonoQ (HR10/10, Pharmacia) operated at a flow rate of 1.5 ml/min; proteins were eluted with a gradient (45 ml) of 20–250 mM NaCl in buffer B. Fractions containing DnaG(434–581) were pooled to yield Fraction IV (64 mg protein), concentrated to 2 ml using an Ultrafree-4 centrifugal filter unit (Millipore), and applied to a column (65 × 2.5 cm) of Sephadex G-50 (Pharmacia) equilibrated with buffer B + 100 mM NaCl. Fractions containing highly purified DnaG(434–581) (24 mg, Fraction V) were pooled and stored frozen in aliquots at −70 °C.

Samples were concentrated by ultrafiltration (as above) and then dialyzed into 10 mM Tris–HCl (pH 7.6), 1 mM dithiothreitol for crystallization, into 10 mM sodium phosphate buffers containing 0.1% NaN_3 for NMR measurements, and into 0.1% formic acid for ESI-MS. The ESI-MS spectrum of DnaG(434–581) gave a measured molecular mass of 16,701 ± 1 Da (cf. the calculated value of 16,701.1 Da for residues 434–581 of DnaG). This indicated that the additional N-terminal methionine residue had been processed in vivo. Protein concentrations in Fractions I–IV were measured by the method of Bradford [23], with bovine serum albumin as standard. Concentrations of pure DnaG(434–581) were determined spectrophotometrically, using a calculated value of ε_{280} of 12,660 $M^{-1}\,cm^{-1}$ [24]. SeMet-DnaG (434–581) and uniformly labeled [^{13}C,^{15}N]DnaG(434–581) were prepared by similar procedures and characterized by ESI-MS. The SeMet derivative had an observed mass of 16,841.4 (cf. a calculated value of 16,841.7 Da). The overall extent of labeling of [^{13}C,^{15}N]DnaG(434–581) was 98.6%.

NMR measurements

^1H NMR spectra were recorded at 600 MHz on a Varian Unity INOVA 600 NMR spectrometer with samples in buffers containing 90% H_2O/10% D_2O. Three separate samples of unlabeled DnaG(434–581) were prepared. Samples I (2.26 mM protein) and II (0.51 mM) were in 10 mM sodium phosphate, pH 7.0, while sample III (0.51 mM) was in 10 mM sodium phosphate, pH 6.05, containing 0.1 M NaCl. Amino acid residues from the mobile N- and C-terminal ends of DnaG(434–581) were identified by a TOCSY spectrum (80 ms mixing time) [25] recorded with sample I at 25 °C. NOESY spectra were recorded to compare the spectral linewidths between samples II and III, at 36 °C. A sample of 0.65 mM [^{13}C,^{15}N]DnaG(434–581) was used to record a ^{15}N-HSQC spectrum under conditions described for sample III, above.

Crystallization

Initial screens (Hampton Crystal Screens I and II) to establish crystallization conditions used vapor diffusion with sitting drops, initially with 1:1 mixtures of stock DnaG(434–581) (7.1 mg/ml) and reservoir buffers. Crystals of native and SeMet-DnaG(434–581) suitable for diffraction experiments were grown by the vapor-diffusion method in hanging drops at 4 °C. Drops containing 1 μl of SeMet-DnaG(434–581) (8.4 mg/ml) and 1 μl of reservoir solution were mixed on siliconized coverslips. The reservoir solution contained 5% (w/v) polyethylene glycol 4000 and 0.2 M ammonium sulfate in 0.1 M sodium acetate buffer, pH 4.6. Reservoir volumes of 1 ml were used. Crystals were frozen following their transfer into the reservoir solution supplemented with glycerol. Crystals were serially transferred to drops containing 5% (v/v) increments in glycerol concentration, up to a maximum of 20% (v/v). Each transfer was followed by a 2 min equilibration period. After the final equilibration step, crystals were frozen by being plunged into liquid nitrogen.

X-ray diffraction analysis

Initial diffraction experiments with crystals of "native" DnaG(434–581) using a rotating anode X-ray generator were unsuccessful. Crystals were subsequently grown at the BioCARS facility, Advanced Photon Source, Argonne, IL, where both native and SeMet-DnaG(434–581) were crystallized. Subsequent diffraction experiments were conducted on beam line 14-ID-B, which was equipped with a cryocooler and Mar CCD-165 area detector, with crystals at 100 K. An X-ray fluorescence scan of energies from 12.6 to 12.7 keV was carried out to ascertain the precise energy of the K absorption-edge for selenium in the SeMet sample. X-ray diffraction data were processed using the HKL package [26].

Results and discussion

Overproduction of DnaG(434–581)

It has been reported that plasmids containing the gene encoding DnaG(434–581) under transcriptional control of a phage T7-promoter cannot be stably maintained in a host strain capable of expression of T7-RNA polymerase [6,13]. We confirmed this observation: introduction of the T7-promoter plasmid pKL1177 into BL21(DE3)/pLysS [27] failed on several occasions to yield viable transformants. This indicates that expression of DnaG(434–581) compromises the viability of *E. coli*, presumably because the protein interacts with the DnaB helicase to titrate the available primase-binding sites.

However, we found that use of tandem phage λ p_R and p_L promoters for expression did permit construction of stably transformed strains. The λ-promoter plasmid pKL1176 (Fig. 1) directs constitutive expression of the temperature-sensitive *cI857* allele of the *c*I gene that encodes the λ repressor. In cells containing this plasmid and growing at 30 °C, the repressor is active and efficiently represses transcription initiated from the two phage promoters. On temperature shift to 42 °C, the thermolabile repressor is inactivated and transcription is enabled [18], leading to expression of DnaG(434–581). Because this mechanism appears to provide more stringent control of transcription than does use of the phage T7-promoter, pKL1176 could be maintained in several different host strains growing at 30 °C, but transformed strains failed to grow at temperatures above 40 °C. This indicated again that expression of DnaG(434–581) is lethal to the host.

Various hosts and growth media were examined to optimize expression of the protein following treatment at 42 °C, as visualized following SDS–PAGE of lysed whole cells. Expression in *E. coli* AN1459 gave three closely spaced protein bands of similar intensity that were not present in cells cultured at 30 °C (data not shown); the smallest had a mobility expected for DnaG(434–581). The production of three different species in this strain appears to be due to suppression of the UGA (ochre) stop codon of *dnaG*. In transcripts from *dnaG* in pKL1176, as in the *E. coli* chromosome, the UGA is followed 18 codons later by another UGA, and this is followed seven codons later by a UAA stop codon. In contrast, use of the suppressor-free strain, BL21(DE3)*recA* [28], as host gave overproduced protein as a single species that accumulated to adequate levels during treatment of cultures for 2–3 h at 42 °C (Fig. 2, lane I).

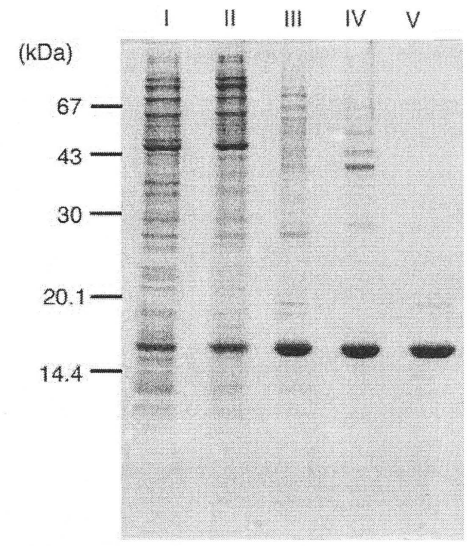


Fig. 2. Overproduction and purification of DnaG(434–581). SDS–PAGE (15%) analysis of samples from successive steps of purification of DnaG(434–581). Lanes are marked I–V, to correspond with fraction numbers given in Materials and methods. A total of 10 μg of protein was loaded in each lane. The gel was stained with Coomassie blue. The mobilities of molecular-weight markers were as indicated (in kDa).

Purification and properties of DnaG(434–581)

The protein was purified by two steps of anion-exchange chromatography, followed by gel filtration, in yields of 4–6 mg/L of culture (Fig. 2). Because of the lethality of DnaG(434–581), we took special care to confirm both the sequence of the gene in the expression plasmid pKL1176 and the molecular weight of the purified protein (by ESI-MS). DnaG(434–581) is a well-behaved protein. It can be concentrated to >35 mg/ml, and is stable on freezing and thawing and for long periods under conditions used for NMR measurements and crystallization. On gel filtration, it eluted in a single peak in a position consistent with it being a monomeric 17 kDa protein (not shown).

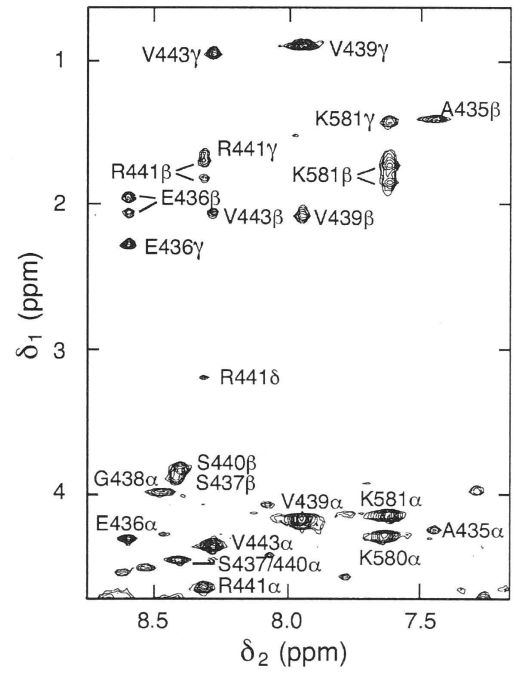


Fig. 3. Spectral region from the TOCSY NMR spectrum recorded with 2.26 mM DnaG(434–581) at pH 7.0 and 25 °C. Crosspeaks with the H^N resonances of mobile amino acid residues are identified with the one-letter amino acid code, their sequence number, and the non-exchangeable proton involved in the crosspeak.

NMR spectroscopy

NMR experiments with several samples of unlabeled DnaG(434–581) were used to comment further on the quaternary structure of the protein, to identify mobile residues at its termini, and to establish conditions for more extensive experiments with isotopically labeled

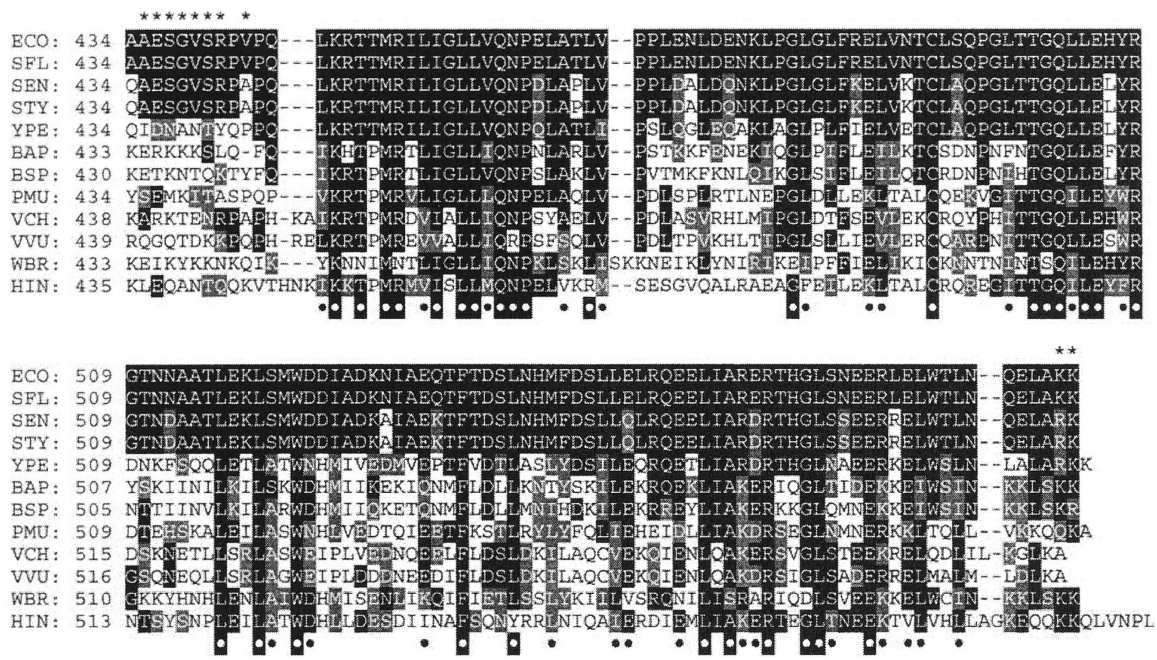


Fig. 4. Alignment of sequences of *E. coli* DnaG(434–581) (*Eco*) with C-terminal regions of primases from *Shigella flexneri* (*Sfl*, GenBank 15803608) and 10 other bacterial species: *Salmonella enterica* (*Sen*, 16761984), *Salmonella typhimurium* (*Sty*, 16766510), *Yersinia pestis* (*Ype*, 16120969), *Buchnera sp.* (*Bsp*, 15616681), *Buchnera aphidicola* (*Bap*, 21672346), *Pasteurella multicocida* (*Pmu*, 15603105), *Vibrio vulnificus* (*Vvu*, 27364090), *Vibrio cholerae* (*Vch*, 15642769), *Haemophilus influenzae* (*Hin*, 16272476), *Wigglesworthia brevipalpis* (*Wbr*, 24323895), and *Xanthomonas campestris* (*Xca*, 21233248). These sequences were chosen as those that showed the highest degree of identity (>30%) to *E. coli* DnaG(434–581) in a BLAST [29] search. Residues identical to those in *Eco* DnaG(434–581) are shown in black, while those that represent conservative substitutions are shown in gray. Dots (below) indicate residues (white) or residue types (black) conserved in at least 11 of the 12 sequences. Asterisks mark mobile residues identified by NMR spectroscopy at the N- and C-termini of DnaG(434–581).

samples. A homonuclear TOCSY spectrum recorded at 25 °C with 2 mM protein in phosphate buffer, pH 7.0 (sample I), indicated that DnaG(434–581) has only a few mobile residues located at the N- and C-termini (Fig. 3). We concluded from this experiment that the structured core of the C-terminal domain of DnaG comprises residues 444–579. It is interesting to note that the last two lysine residues are not structured.

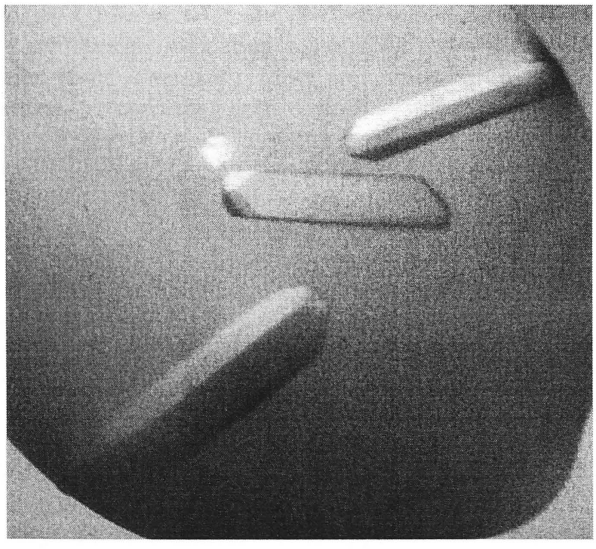


Fig. 5. Crystals of DnaG(434–581). Typical dimensions of crystals were 0.5 × 0.1 × 0.1 mm.

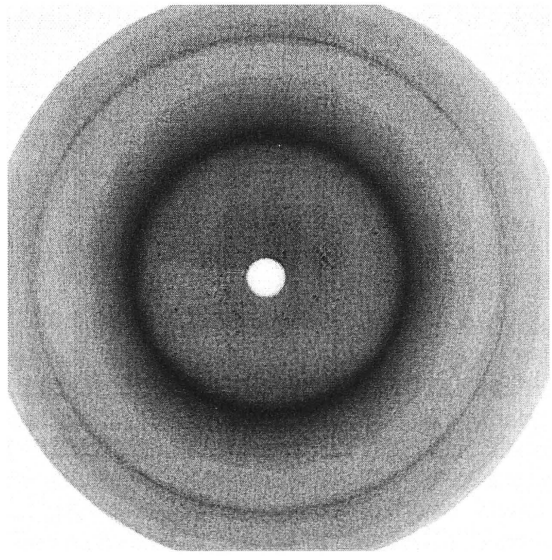


Fig. 6. X-ray diffraction image obtained from a crystal of selenomethionine-substituted DnaG(434–581). The diffraction limit occurs at 2.8 Å.

Table 1
Data collection statistics for crystals of selenomethionine-substituted DnaG(434–581)

Crystal number	1	1	2
Wavelength (Å)	0.979862 (peak λ_1)	0.98004 (inflection λ_2)	0.9776 (remote λ_3)
Space group	$I4_122$	$I4_122$	$I4_122$
a (Å)	142.2	142.2	141.7
c (Å)	192.1	192.1	191.1
Resolution range (Å)	100–2.8 (2.9–2.8)	100–3.1 (3.21–3.2)	100–2.9 (3.00–2.9)
Numbers of reflections			
Measured	135,951	101,900	126,466
Unique	24,195	17,992	21,472
Multiplicity	5.6	5.7	5.9
R_{merge}[a] (%)	9.9 (68.9)	10.6 (69.1)	9.4 (56.8)
Completeness (%)	98.7 (100)	98.7 (100)	98.9 (100)
$I/\sigma(I)$	12.89 (1.8)	11.2 (2.9)	14.3 (3.14)

[a] $R_{merge} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ is the ith measurement of independent intensity.

Even though a pair of basic residues is generally well conserved at the C terminus of bacterial primases (Fig. 4), the substitution of Lys581 with Ala had no detectable effect on interaction of DnaG(434–581) with the DnaB helicase, while the Lys580 to Ala substitution in full-length primase reduced DnaB-dependent priming activity about twofold [3]. That residues 434–443 are not well structured is consistent with a comparison of sequences of primases most closely related to the *E. coli* protein (Fig. 4); sequences in this region are poorly conserved relative to those in the structured core of the domain. Given that residues 429–433 at the C-terminus of the central domain were also disordered in crystals [7], this suggests that DnaG has a flexible linker of about 15 residues that connect the two domains. This may be required to accommodate movement of the domains as they function in a primosome at a replication fork.

A NOESY spectrum (recorded at 36 °C) of a 0.5 mM solution of DnaG(434–581) in phosphate buffer, pH 7.0 (sample II), showed broad lines for most of the crosspeaks, indicating aggregation of the protein under these conditions (data not shown). However, spectra recorded with both unlabeled and [^{13}C,^{15}N]DnaG(434–581) at lower pH (phosphate buffer, pH 6.05) in the presence of 0.1 M NaCl (conditions of sample III) showed much narrower linewidths characteristic of a monomeric protein. DnaG(434–581) is expected to be a monomeric protein at high concentrations in solution, since full-length primase is monomeric under all conditions that have been examined [11]. Less than 10% of the labeled sample of protein precipitated during more than a week in the NMR spectrometer at 36 °C.

X-ray crystallography

Crystals of DnaG(434–581) appeared as tetragonal prisms after one day and grew to a maximum size of about 1 mm in the longest dimension after two days (Fig. 5). The crystals were very sensitive. Since moving them from the cold room led to their rapid disintegration, all manipulations were carried out at 4 °C.

Although most diffracted poorly, a few crystals of SeMet-DnaG(434–581) diffracted to beyond 3 Å at 100 K, using synchrotron radiation (Fig. 6). One of these was used for an X-ray fluorescence scan, as well as for data collection experiments at different wavelengths. Based on the results (data not shown), three wavelengths were chosen, one at the peak, one at the inflection point, and one remote from these (Table 1). The crystals deteriorated in the beam, and it was not possible to collect three complete data sets from a single crystal. Therefore, a second crystal, selected from the same drop as the first and prepared identically, was used to collect the final data set at the remote wavelength. The data were successfully indexed in an I-centered tetragonal space group in all cases. Analysis of reflections of the type 00L and H00 revealed the space group to be $I4_122$. Statistics for data collection are shown in Table 1.

Studies are underway to phase and solve the crystal structure of DnaG(434–581), and to assign resonances in its NMR spectra to permit detailed examination of its interaction with the DnaB helicase.

Acknowledgments

We are grateful to the Australian Research Council for research grant support. A.J.O. was a recipient of an Australian Postdoctoral Fellowship and K.L. was supported by a scholarship from the Deutscher Akademischer Austauschdienst. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under Grant No. RR07707.

References

[1] A. Kornberg, T.A. Baker, DNA Replication, second ed., Freeman, New York, 1991.

[2] R. McMacken, A. Kornberg, A multienzyme system for priming the replication of ϕX174 viral DNA, J. Biol. Chem. 253 (1978) 3313–3319.

[3] K. Tougu, K.J. Marians, The interaction between helicase and primase sets the replication fork clock, J. Biol. Chem. 271 (1996) 21398–21405.

[4] L.E. Bird, H. Pan, P. Soultanas, D.B. Wigley, Mapping protein–protein interactions within a stable complex of DNA primase and DnaB helicase from *Bacillus stearothermophilus*, Biochemistry 39 (2000) 171–182.

[5] T.V. Ilyina, A.E. Gorbalenya, E.V. Koonin, Organization and evolution of bacterial and bacteriophage primase–helicase systems, J. Mol. Evol. 34 (1992) 351–357.

[6] K. Tougu, H. Peng, K.J. Marians, Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork, J. Biol. Chem. 269 (1994) 4675–4682.

[7] J.L. Keck, D.D. Roche, A.S. Lynch, J.M. Berger, Structure of the RNA polymerase domain of *E. coli* primase, Science 287 (2000) 2482–2486.

[8] M. Podobnik, P. McInerney, M. O'Donnell, J. Kuriyan, A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases, J. Mol. Biol. 300 (2000) 353–362.

[9] H. Pan, B.D. Wigley, Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase, Structure 8 (2000) 231–239.

[10] T. Kusakabe, C.C. Richardson, The role of the zinc motif in sequence recognition by DNA primases, J. Biol. Chem. 271 (1996) 19563–19570.

[11] N.P.J. Stamford, P.E. Lilley, N.E. Dixon, Enriched sources of *Escherichia coli* replication proteins. The *dnaG* primase is a zinc metalloprotein, Biochim. Biophys. Acta 1132 (1992) 17–25.

[12] M. Kato, T. Ito, G. Wagner, C.C. Richardson, T. Ellenberger, Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis, Mol. Cell 11 (2003) 1349–1360.

[13] K. Tougu, K.J. Marians, The extreme C terminus of primase is required for interaction with DnaB at the replication fork, J. Biol. Chem. 271 (1996) 21391–21397.

[14] P. Chang, K.J. Marians, Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork, J. Biol. Chem. 275 (2000) 26187–26195.

[15] J. Weigelt, S.E. Brown, C.S. Miles, N.E. Dixon, G. Otting, NMR structure of the N-terminal domain of *E. coli* DnaB helicase: implications for structure rearrangements in the helicase hexamer, Structure 7 (1999) 681–690.

[16] D. Fass, C.E. Bogden, J.M. Berger, Crystal structure of the N-terminal domain of the DnaB hexameric helicase, Structure 7 (1999) 691–698.

[17] S.G. Vasudevan, W.L.F. Armarego, D.C. Shaw, P.E. Lilley, N.E. Dixon, R.K. Poole, Isolation and nucleotide sequence of the *hmp* gene that encodes a haemoglobin-like protein in *Escherichia coli* K-12, Mol. Gen. Genet. 226 (1991) 49–58.

[18] C.M. Elvin, P.R. Thompson, M.E. Argall, P. Hendry, N.P.J. Stamford, P.E. Lilley, N.E. Dixon, Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*, Gene 87 (1990) 123–126.

[19] C.A. Love, P.E. Lilley, N.E. Dixon, Stable high-copy-number bacteriophage λ promoter vectors for overproduction of proteins in *Escherichia coli*, Gene 176 (1996) 49–53.

[20] C. Neylon, S.E. Brown, A.V. Kralicek, C.S. Miles, C.A. Love, N.E. Dixon, Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance, Biochemistry 39 (2000) 11989–11999.

[21] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning. A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[22] G.D. van Duyne, R.F. Standaert, P.A. Karplus, S.L. Schreiber, J. Clardy, Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin, J. Mol. Biol. 229 (1993) 105–124.

[23] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.

[24] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326.

[25] J. Briand, R.R. Ernst, Computer-optimized homonuclear TOCSY experiments with suppression of cross relaxation, Chem. Phys. Lett. 185 (1991) 276–285.

[26] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol. 276 (1997) 307–326.

[27] F.W. Studier, A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185 (1990) 60–89.

[28] N.K. Williams, P. Prosselkov, E. Liepinsh, I. Line, A. Sharipo, D.R. Littler, P.M.G. Curmi, G. Otting, N.E. Dixon, In vivo protein cyclization promoted by a circularly permuted *Synechocystis* sp. PCC6803 DnaB mini-intein, J. Biol. Chem. 277 (2002) 7790–7798.

[29] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.